

**FIRST JUDICIAL DISTRICT COURT
COUNTY OF SANTA FE
STATE OF NEW MEXICO**

STATE OF NEW MEXICO, EX REL.,)
RAÚL TORREZ, ATTORNEY GENERAL,)
)
Plaintiff,)
)
v.)
)
META PLATFORMS, INC.; INSTAGRAM, LLC;)
META PAYMENTS, INC.; META PLATFORMS)
TECHNOLOGIES, LLC; and)
MARK ZUCKERBERG,)
)
Defendants.)

NO. _____

PLAINTIFF'S COMPLAINT FOR ABATEMENT AND CIVIL PENALTIES
AND DEMAND FOR JURY TRIAL

**NOTE: THIS COMPLAINT, ESPECIALLY AT PAGES 25-85 CONTAINS REDACTED
IMAGES OF CHILD SEXUAL EXPLOITATION AND OTHER SEXUALLY EXPLICIT
IMAGES, AS WELL AS IMAGES OF SELF HARM, WHICH MAY BE DISTURBING.**

TABLE OF CONTENTS

I.	INTRODUCTION.....	1
II.	PARTIES	3
III.	JURISDICTION, VENUE, AND JURY DEMAND	6
IV.	META ENGAGES IN TRADE OR COMMERCE WITHIN NEW MEXICO....	8
V.	META’S DESIGN OF INSTAGRAM AND FACEBOOK.....	15
VI.	META’S PLATFORMS ENABLED AND FAILED TO PREVENT SEXUAL EXPLOITATION MATERIAL FROM AFFECTING NEW MEXICO RESIDENTS.....	21
A.	DEFINITION OF HUMAN TRAFFICKING.....	21
B.	META STEERED AND CONNECTED USERS – INCLUDING CHILDREN – TO SEXUALLY EXPLICIT, EXPLOITATIVE AND CHILD SEX ABUSE MATERIALS AND FACILITATED HUMAN TRAFFICKING WITHIN OR AFFECTING NEW MEXICO.....	23
i.	Case Study 1: Rosalind Cereceres	37
ii.	Case Study 2: Issa Bee	44
iii.	Case Study 3: Sunny Paxton	60
iv.	Case Study 4: BobbiFun13.....	64
v.	Case Study 5: Taya Neils.....	65
vi.	Case Study 6: Sophia Keys.....	74
VII.	THE HARMFUL CONTENT ON META’S PLATFORMS REMAINS AND IS PROLIFERATED BY META’S ALGORITHMS.....	85
VIII.	HISTORIC NATURE OF META’S UPA VIOLATIONS DISCOVERED BY OTHER INVESTIGATIONS	88
IX.	META HAS LONG BEEN ON NOTICE OF, BUT FAILED TO ADDRESS, COMMERICAL SEXUAL EXPLOITATION AND TRAFFICKING OF CHILDREN THROUGH SOCIAL MEDIA AND ITS PLATFORMS SPECIFICALLY.....	100
X.	META PROFITS FROM THE ALL-TOO-COMMON DISTRIBUTION AND SOLICITATION OF CSAM	104
XI.	CHILDREN IN NEW MEXICO HAVE FALLEN VICTIM TO DISTRIBUTION AND SOLICITATION OF CSAM OR CSEC ON META’S PLATFORMS	109
XII.	META WAS, AND IS, AWARE THAT ITS PLATFORMS ARE BEING USED TO TARGET, GROOM, SEXUALLY EXPLOIT AND TRAFFIC CHILDREN	113

XIII.	META’S BUSINESS MODEL TARGETS YOUNG USERS SPECIFICALLY WITH FEATURES DESIGNED TO ENTICE AND ADDICT YOUTH.....	132
A.	ENGAGEMENT-BASED FEEDS AND THE ALGORITHMS.....	137
B.	INFINITE SCROLL.....	141
C.	PUSH NOTIFICATIONS	143
D.	EPHEMERAL CONTENT	145
E.	VIDEO CONTENT.....	146
F.	ADDITIONAL DESIGN CHOICES.....	148
XIV.	META WAS ACUTELY AWARE OF THE HARM TO YOUTH WELL-BEING RESULTING FROM ITS DESIGN CHOICES, BUT FAILED TO DEVOTE SUFFICIENT RESOURCES TO ADEQUATELY ADDRESS THE HARM TO YOUTH	151
XV.	DESPITE KNOWLEDGE OF THE HARMS CAUSED BY ITS PLATFORMS, META REPEATEDLY REJECTED EFFORTS TO IMPROVE ITS SYSTEMS	167
XVI.	META’S PLATFORM DESIGN CAUSED MENTAL HEALTH HARM TO YOUNG USERS, INCLUDING USERS IN NEW MEXICO.....	173
XVII.	META MADE MISLEADING AND DECEPTIVE CLAIMS IN PUBLIC STATEMENTS THAT ITS PLATFORM WAS SAFE OR THAT IT WAS ADDRESSING PROBLEMATIC CONTENT	177
XVIII.	ADDITIONAL ALLEGATIONS RELATING TO MARK ZUCKERBERG ..	186
XIX.	CAUSES OF ACTION	200
XX.	PRAYER FOR RELIEF.....	222

I. INTRODUCTION

1. Meta and its CEO tell the public that Meta's social media platforms are safe and good for kids. The reality is far different. Meta knowingly exposes children to the twin dangers of sexual exploitation and mental health harm. Meta's conduct has turned New Mexico children who are on its platforms into victims. Meta's motive for doing so is profit. This action seeks to make social media safer for New Mexico's children by holding Meta accountable for conduct that violates the New Mexico Unfair Practices Act and creates a public nuisance.

2. Meta's platforms Facebook and Instagram are a breeding ground for predators who target children for human trafficking, the distribution of sexual images, grooming, and solicitation. Teens and preteens can easily register for unrestricted accounts because of a lack of age verification. When they do, Meta directs harmful and inappropriate material at them. It allows unconnected adults to have unfettered access to them, which those adults use for grooming and solicitation. And Meta's platforms do this even though Meta has the capability of both determining that these users are minors and providing warnings or other protections against material that is not only harmful to minors but poses substantial dangers of solicitation and trafficking. For years, Meta has been on notice from both external and internal sources of the sexual exploitation dangers its platforms present for children but has nonetheless failed to stem the tide of damaging sexual material and sexual propositions delivered to children. In short, Meta has allowed Facebook and Instagram to become a marketplace for predators in search of children upon whom to prey. Meta's conduct is not only unacceptable; it is unlawful. This action seeks to force Meta to institute protections for children because it refuses to do so voluntarily.

3. Simultaneously with its knowing failure to curb the sexual exploitation of children on its platforms, Meta targeted the age-based vulnerabilities of children by adopting algorithms and platform designs that are addictive to young users. Meta knowingly sought to maximize teen

engagement on its platforms. It chose to implement features such as engagement-based feeds, infinite scroll, push notifications, ephemeral content, and auto play video designed to increase the amount of time young users spend on its platforms while inhibiting the ability of those users to self-regulate. Meta's platforms are the social media equivalent of an addictive drug from which young users cannot break free. Meta knew that these design features fostered addiction, anxiety, depression, self-harm, and suicide among teens and preteens. But Meta and its CEO rejected repeated internal proposals, and external pressures, to implement protections against youth mental health harm. Further, Meta selected a metric by which it measures conduct violative of its Community Standards policies that it knows to grossly underreport harmful material on its platforms, and Meta uses this metric to make misrepresentations about the safety of its platforms for young users.

4. Meta profits from its exposure of young users to harmful material and its refusal to implement design features that would protect children from sexual exploitation and mental health harm. It does so not by charging children for accessing its platforms but instead by monetizing, in the form of targeted advertising, the data that Meta gathers about its young users and their usage. Meta's "targeted" advertising program allows advertisers to direct advertisements to consumers more precisely than would otherwise be possible using traditional media. This arrangement has proved particularly lucrative for Meta. The company reported more than *\$116 billion* in revenue in 2022, and *\$117 billion* the year before. As Meta's financials confirm, all or substantially all of this revenue is attributable to advertising and enhanced by its user-data-driven ability to target advertising.

5. Meta's platforms must maintain massive user bases in order to generate its target revenue. Meta must not only attract new users year over year, but it must ensure that existing users

remain on its platforms. If users leave Facebook or Instagram, if new users refuse to join altogether, or if these users spend less time on its platforms, Meta's revenues will suffer as it would have less private data, and fewer users, to sell. As Meta warns investors in its annual SEC filings, "If we fail to retain existing users or add new users, or if our users decrease their level of engagement with our products, our revenue, financial results, and business may be significantly harmed."¹

6. Meta's business model of profit over child safety and business practices of misrepresenting the amount of dangerous material and conduct to which its platforms expose children violates New Mexico law. Meta should be held accountable for the harms it has inflicted on New Mexico's children and be required to make its platforms as safe for children as the law requires.

II. PARTIES

7. This action is brought by the State of New Mexico in its sovereign capacity by and through Raúl Torrez, the Attorney General of the State of New Mexico. The Attorney General acts pursuant to his authority under, *inter alia*, NMSA 1978, Sections 8-5-1 to -17 (1933, as amended through 1999); the New Mexico Unfair Practices Act, NMSA 1978, Sections 57-12-1 to -26 (1967, as amended through 2009); and NMSA 1978, Sections 30-8-1, 30-8-8 (1963).

8. Defendant Meta Platforms, Inc., d/b/a Meta ("Meta"), and formerly known as Facebook, Inc. and TheFacebook, Inc., is a multinational technology company incorporated in Delaware, with a principal place of business in Menlo Park, California. Defendant Meta Platforms, Inc. is registered to do business in New Mexico as a Foreign Profit Corporation.

9. Meta, through itself and/or its subsidiaries, develops, designs, markets, and operates social media platforms and services including Facebook, Instagram, Messenger, and

¹ Meta 2023 10-K at p. 15.

WhatsApp (collectively, Meta’s “platforms”). As a result of acquisitions such as Instagram, Meta has come to dominate the market of social media products and apps, becoming the largest social media company in the world.² As of November 2023, Meta has a market capitalization of \$864.27 billion.

10. For purposes of this Complaint, Meta includes its wholly-owned subsidiaries Instagram, LLC; Facebook Holdings, LLC; Facebook Operations, LLC; Meta Payments Inc.; Meta Platforms Technologies, LLC; and Siculus, Inc.

11. Defendant Instagram, LLC (“Instagram”), offers an online social media networking platform that enables users to post and share images and videos with others, as well as interact with other users. Instagram is a limited liability company incorporated in Delaware with its principal place of business in Menlo Park, California. Meta purchased Instagram on April 9, 2012 and Instagram, since then, is a wholly-owned subsidiary of Meta. Instagram currently has over two billion monthly active users worldwide,³ and an estimated 62% of U.S. teens are on the platform.⁴ Although Defendant Instagram, LLC offers services to and conducts business with New Mexico residents as explained below, Defendant Instagram, LLC is not registered to do business in New Mexico.

12. Defendant Meta Payments Inc. is incorporated in the State of Florida and shares its principal place of business in Menlo Park, California, with Meta. Meta Payments Inc. processes payments made through Meta’s social media platforms. Meta directly owns Meta Payments Inc.,

² Andrea Murphy, et al., *The Global 2000*, FORBES (June 8, 2023), <https://www.forbes.com/lists/global2000/>; see also Stacy Jo Dixon, *Global Social Networks Ranked by Number of Users 2023*, STATISTA (Sept. 22, 2023), <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>.

³ Meta Platforms, Inc. Third Quarter 2022 Results Conference Call Transcript (Oct. 26, 2022), available at: https://s21.q4cdn.com/399680738/files/doc_financials/2022/q3/Meta-Q3-2022-Earnings-Call-Transcript.pdf

⁴ Emily A. Vogels, Risa Gelles-Watnick & Navid Massarat, *Teens, Social Media and Technology 2022*, Pew Research Center (Aug. 10, 2022), <https://www.pewresearch.org/internet/2022/08/10/teens-social-media-and-technology-2022/>

its subsidiary. Defendant Meta Payments Inc. is registered to do business in New Mexico as a Foreign Profit Corporation.

13. Defendant Meta Platforms Technologies, LLC is a Delaware limited liability company and shares its principal place of business in Menlo Park, California, with Meta. Previously known as Facebook Technologies, LLC, Meta Platforms Technologies, LLC has absorbed Meta's Oculus business segment, which it acquired in 2014. Meta Platforms Technologies, LLC develops Meta's virtual reality technology. Defendant Meta is the sole member or manager of Meta Platforms Technologies, LLC. Defendant Meta Platforms Technologies, LLC is registered to do business in New Mexico as a Foreign Limited Liability Company.

14. At all relevant times, Meta, including through its subsidiaries and executives, collectively directed, controlled, had the authority to control, or participated in all aspects of the strategy, operation, planning, management, policies, design, and development of its social media platforms, including in the acts and practices set forth in this Complaint.

15. As detailed in the allegations below, Meta is engaging, has engaged, and continues to engage in unfair, deceptive, unconscionable, and unlawful activity in New Mexico. Meta has conducted this activity on its own and/or through its subsidiaries over which it exercises complete control and dominion, and over which Meta's executive officers and directors have direct oversight. Because Meta and its subsidiaries operate as a common enterprise, each of them is jointly and severally liable for the acts and practices alleged below. All references to "Meta" in this Complaint shall refer to and are meant to include both Meta and its above-named subsidiaries and platforms, including Facebook, Instagram, and WhatsApp.

16. According to his official Meta biography, Mark Zuckerberg is the "founder, chairman and CEO of Meta, which he originally founded as Facebook in 2004. He is responsible

for setting the overall direction and product strategy for the company,”⁵ and “has control over key decision making as a result of his control of a majority of the voting power of our outstanding capital stock.”⁶ He has acted as lead spokesperson for Meta, speaking through public statements, his Facebook account, and through Congressional testimony. Forbes lists Zuckerberg as the eighth wealthiest individual in the world, with a net worth estimated at \$106 billion. He personally owns 13% of Meta, which comprises the majority of his wealth. Zuckerberg is sued in his individual capacity, for his own actions, inactions, statements and omissions.

III. JURISDICTION, VENUE, AND JURY DEMAND

17. Subject matter jurisdiction for this case is conferred upon this Court pursuant to, *inter alia*, Article VI, Section 13 of the New Mexico Constitution.

18. This Court has personal jurisdiction over Defendants because Defendants do business in New Mexico and/or have the requisite minimum contacts with New Mexico necessary to constitutionally permit the Court to exercise jurisdiction with such jurisdiction also within the contemplation of the New Mexico “long arm” statute, NMSA 1978, Section 38-1-16 (1971).

19. Specifically, the Court has personal jurisdiction over Defendant Zuckerberg because he actively and personally participated in creating, directing, delivering, or approving deceptive, unfair, and unconscionable conduct that was delivered to and/or affected New Mexico. Zuckerberg made decisions and directed actions that caused, extended, and failed to mitigate harms that Meta’s social media platforms caused, and continue to cause, children and other users of its products—and personally and directly benefitted from those decisions and actions. More

⁵ <https://about.meta.com/media-gallery/executives/mark-zuckerberg/> (last visited Nov. 29, 2023).

⁶ <https://www.sec.gov/Archives/edgar/data/1326801/000132680123000013/meta-20221231.htm> (last visited Nov. 29, 2023).

generally, Zuckerberg controlled, directed, approved, and/or ratified the harmful and unlawful conduct described in this Complaint.

20. The State brings this action exclusively under the law of the State of New Mexico. No federal claims are being asserted, and to the extent that any claim or factual assertion set forth herein may be construed to have stated any claim for relief arising under federal law, such claim is expressly disavowed and disclaimed by the State. The State's citation to federal statutes is only to underscore public policy and standards that inform the State's claims that Defendants' conduct is deceptive, unfair, and unconscionable and constitutes a public nuisance under New Mexico law and are not alleged as independent claims or causes of action.

21. The State does not seek to hold Meta liable as the publisher or speaker of any of the content described herein. Rather, the State asserts claims against Meta are based upon Meta's deceptive, unfair, unconscionable, unreasonable, and unlawful conduct in designing and maintaining its products (Instagram and Facebook, in particular) in such a manner so as to cause known harms to its users; and making deceptive statements concerning Meta's conduct, platforms and policies that, in fact, constituted misrepresentations or contained material omissions concerning the content existing on Meta's platforms and Meta's dedication and/or efforts to combat that conduct. Further, to the extent the allegations are construed to hold Meta liable as the publisher or speaker of content on its platforms, such claims fall within exceptions to such liability.

22. Venue is proper in this Court pursuant to NMSA 1978, Section 38-3-1 (1988), because the Office of the Attorney General and the seat of the State Government are situated in the City and County of Santa Fe, State of New Mexico, and the claims for relief asserted herein arose in large part in the City and County of Santa Fe, New Mexico.

23. Pursuant to Rules 1-038 (A) and (B)(1) NMRA, Plaintiff hereby demands trial by jury of twelve persons. As a State agency, the Attorney General's office is exempt from paying a jury demand fee.

IV. META ENGAGES IN TRADE OR COMMERCE WITHIN NEW MEXICO

24. As required by the UPA, Meta has engaged in trade or commerce within New Mexico. Meta designed its platforms to monetize its users' private data as a form of currency that it uses to secure revenue from targeted advertising. Meta allows its platforms to operate in a manner that cultivates an atmosphere for the creation and proliferation of harmful content including CSAM and sex trafficking and designed its platforms in a manner that fosters mental health harm and self-harm among teens and preteens.

25. Meta operates at least three online platforms that are relevant to this action: Facebook, Instagram, and WhatsApp. All three platforms are offered to and used by New Mexico citizens and/or citizens of other states traveling in or visiting New Mexico in the course of Meta's commercial activities. The platforms do not operate on a state-specific basis; nor do they employ electronic geographical boundaries that restrict usage in New Mexico.

26. Meta does not charge users a monetary fee to use its products; instead, it monetizes consumers' private data by actively harvesting it and using it to sell lucrative advertising. Meta generates "substantially all of [its] revenue from selling advertising placements on our family of apps to marketers."⁷ The essential part of Meta's advertising sales is the data Meta gathers from its users. Meta discloses in its financial statements that "[t]he size of our user base and our users' level of engagement across our products are critical to our success. Our financial performance has been and will continue to be significantly determined by our success in adding, retaining, and

⁷ Meta Form 10-K at 7.

engaging users of our products that deliver ad impressions, particularly for Facebook and Instagram . . . [D]eclines in the size of our active user base may adversely impact our ability to deliver ad impressions and, in turn, our financial performance.”⁸ Retaining users and maintaining or increasing their level of engagement is thus a key focus for Meta, and Meta’s own securities filings confirm that a decline in users or user engagement would result in a decline in revenue.

27. The phrase “ad impressions” generally refers to the number of “views” a particular advertisement receives on one of Meta’s platforms. Meta maximizes the number of “ad impressions” by collecting data from their users and then monetizes that data by using it, in the aggregate, to target advertisements to demographics or individuals with characteristics that advertisers find appealing.

28. Meta’s Terms of Service for its Facebook and Instagram products make that *quid pro quo* explicit: instead of paying with currency, users of Facebook and Instagram provide their demographic and usage data to Meta, which, in turn, Meta uses on an aggregate basis to entice advertisers to pay Meta to place advertisements on Meta’s platforms that are targeted to specific audiences. As Meta advises in its Facebook Terms of Service, “we don’t charge you to use Facebook or the other products and services covered by these Terms . . . [i]nstead, businesses and organizations, and other persons pay us to show you ads for their products and services . . . [w]e use your personal data to help determine which personalized ads to show you.”

29. Meta’s Instagram Terms of Use contain a similar disclosure, confirming the same *quid pro quo*, noting that “[i]nstead of paying to use Instagram . . . you acknowledge that we can show you ads that businesses and organizations pay us to promote . . . We use your personal data,

⁸ *Id.* at 15.

such as information about your activity and interests, to show you ads that are more relevant to you” and provide data on the ads’ performance to advertisers.

30. Meta executives have confirmed the importance of collecting users’ data in order to deliver targeted advertisements. In a January 30, 2019 earnings call CFO David Wehner stated: “In terms of our ability to continue to grow the advertising business, it’s about working to develop the best – the best products we can to enable advertisers to achieve their end business results. Targeting obviously is very important in that.”⁹

31. Meta advertises its ability to target advertisements to potential advertisers, including on its website where it promotes that advertisers are able to reach the people that conform to the specific “traits . . . interests, gender, or location” of the demographic the advertisers seek.¹⁰

32. Potential advertisers then are attracted to Meta because of Meta’s ability to target advertising to selected users/demographics and because of Meta’s ability to provide metrics regarding advertising performance. In this way, Meta’s advertising platform is unique from traditional advertising, in which advertisers place their advertisements on billboards, in magazines or on television with broad distribution that includes both consumers who may be interested in the product and consumers who may not be interested. Thus, Meta’s ability to collect and aggregate data and to enable advertisers to incorporate the collected data into an advertising strategy are a key selling point for Meta.

33. However, companies that seek to advertise on Meta “do not have long-term advertising commitments.”¹¹ They “will not continue to do business with [Meta], or they will reduce the budgets they are willing to commit to [Meta], if [Meta] do[es] not deliver ads in an

⁹ https://s21.q4cdn.com/399680738/files/doc_financials/2018/Q4/Q4-2018-earnings-call-transcript.pdf at 17 (edited for clarity) (last visited Nov. 29, 2023).

¹⁰ <https://www.facebook.com/business/ads/ad-targeting> (last visited Nov. 29, 2023).

¹¹ Meta 2023 Form 10-K at 17.

effective manner” or “if they do not believe that their investment in advertising with us will generate a competitive return relative to other alternatives.”¹² Thus, Meta must attract and retain a large user base and ensure that they are online to receive advertisements in order to generate revenue.

34. Happily for Meta, this *quid pro quo* has proved lucrative. The company reported \$34 billion in revenue in the third quarter of 2023.¹³ As of September 2023, according to Meta, an average of 3.14 billion people used one of Meta’s products on a daily basis (“Daily Average Users” or “DAUs”) and 3.96 billion people used one of Meta’s products on a monthly basis (“Monthly Average Users” or “MAUs”).

35. Meta’s platforms engage in other forms of commerce in New Mexico besides advertising. For example, Facebook includes a feature called “Facebook Marketplace,” which facilitates the sale of goods and services via the Facebook platform by permitting users to post the digital equivalent of “classified advertisements” on Facebook. Meta charges and collects a fee when items are sold on its Marketplace, including in New Mexico. Additionally, Meta offers users, including users in New Mexico, the opportunity to monetize their account(s) in order to sell subscriptions or permit advertisements to be placed on their platforms. Indeed, Facebook publishes a set of “Partner Monetization Policies” and “Content Monetization Policies” that specify the requirements for such accounts.

36. Additionally, Facebook users who meet certain criteria are eligible to participate in “Facebook Stars,” a program whereby the user receives “stars” or “gifts” that other users purchase directly from Meta. In order to qualify for this program, a user must have 500 followers for at least

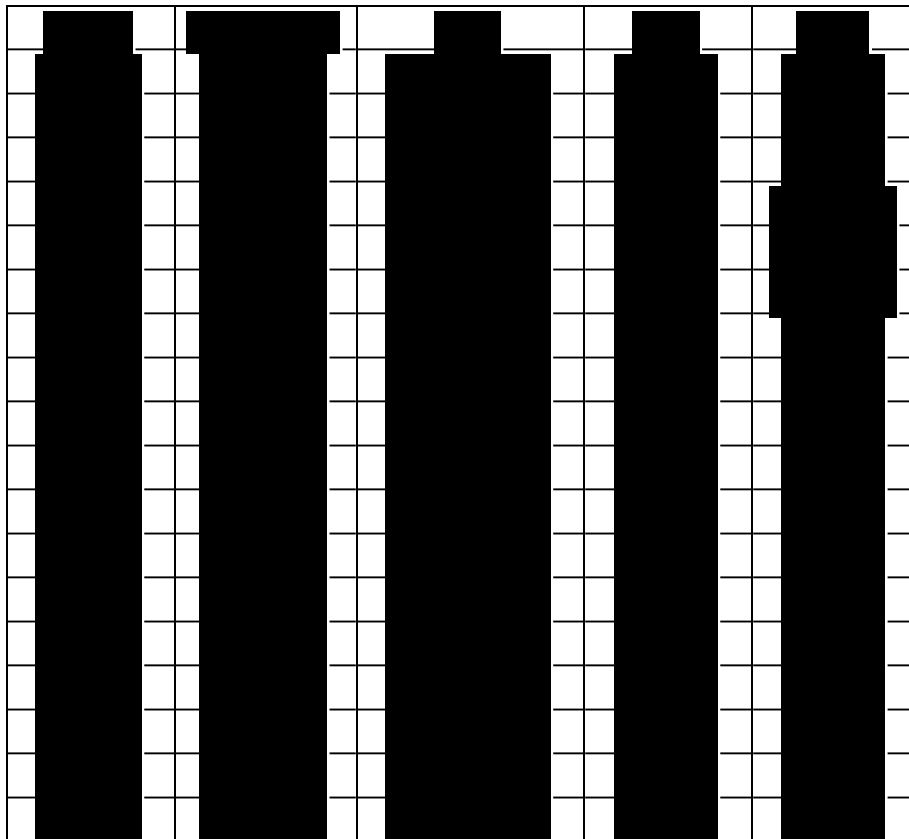
¹² *Id.*

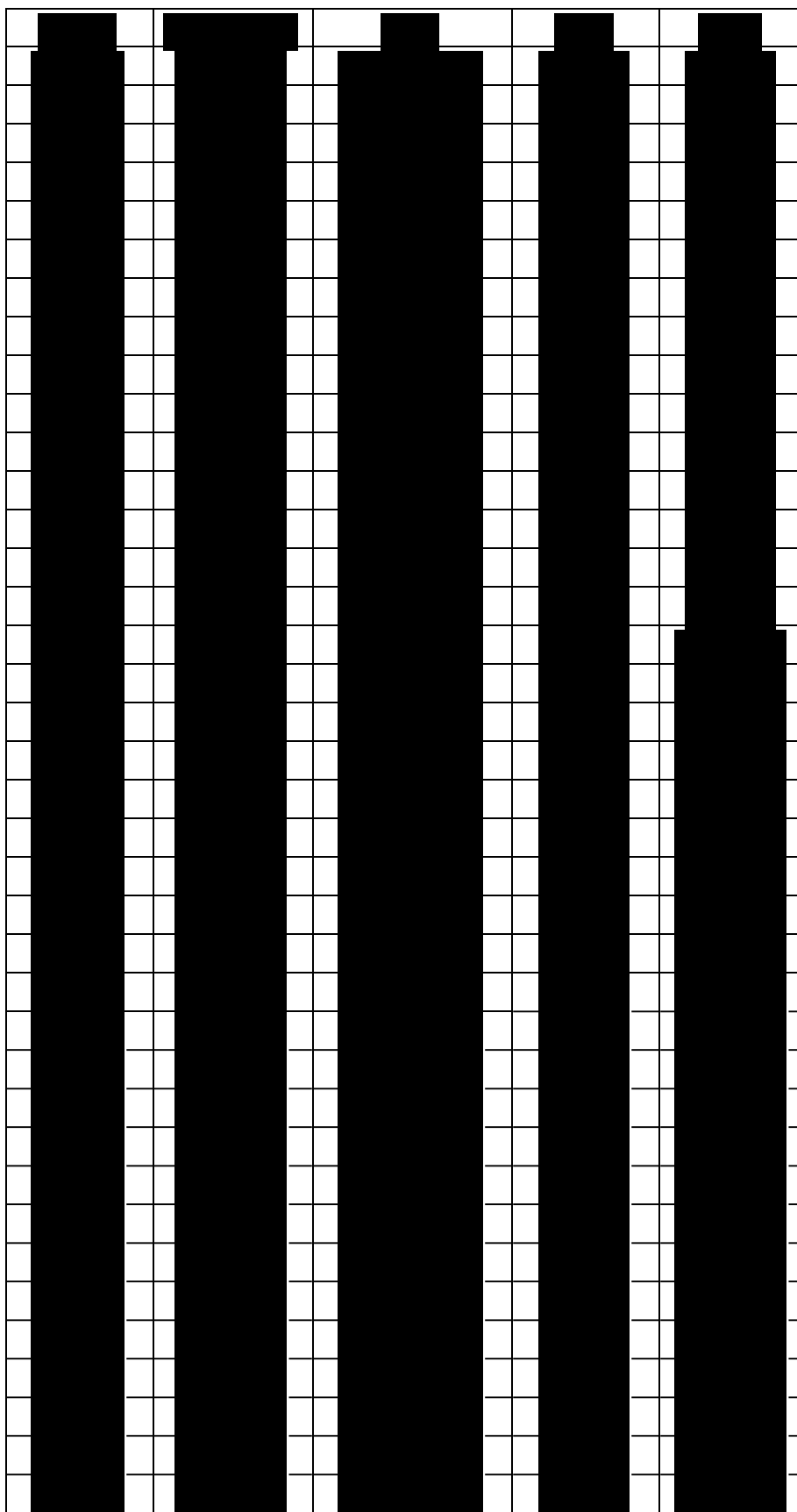
¹³ https://s21.q4cdn.com/399680738/files/doc_earnings/2023/q3/earnings-result/Meta-09-30-2022-Exhibit-99-1-FINAL.pdf

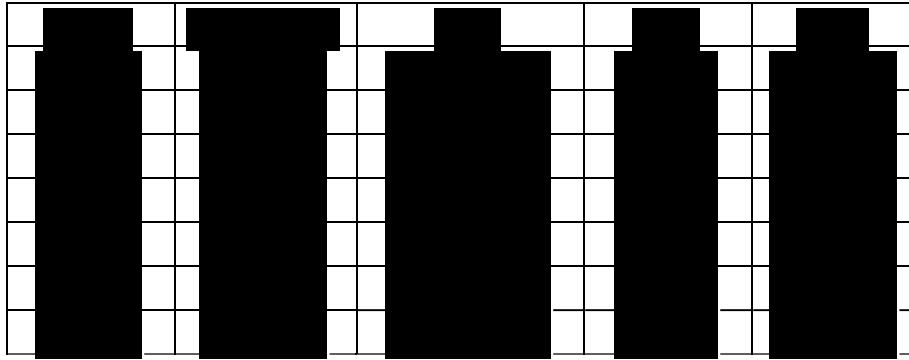
30 consecutive days and be over 18 years old, among other criteria. Meta advertises the opportunity for users to “earn money from Facebook Stars” if they have a “professional” account. “Professional” accounts, like businesses, have the ability to monetize their content through the placement of advertisements on Reels that the user posts, through the placement of advertisements that play directly before or after Reels that the user posts, or by selling subscriptions.

37. In addition to this “Professional” mode, Meta offers users the ability to purchase advertisements in order to promote their own Instagram profile, in addition to traditional advertisements selling goods and services.

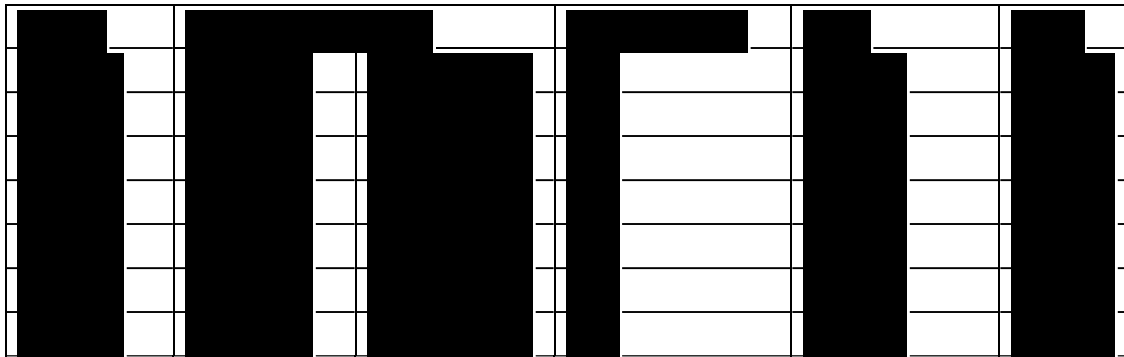
38. A significant number of New Mexico residents use Meta’s products. Between February 2018 and April 2023, Instagram DAUs averaged 609,297, with more than 760,000 New Mexico DAUs reported in March 2023. Instagram MAUs in New Mexico were even higher, averaging 958, 801, with a high of 1,195,544 MAUs in January 2023:







39. Meta does not appear to have tracked or recorded the ages of its Instagram users until October 2022 because statistics regarding earlier periods were not produced to the Attorney General. However, information produced to the Attorney General confirms that significant numbers of users that Meta classified as “teens” in New Mexico used and continued to use Instagram. Between October 2022 and April 2023, Instagram DAUs averaged 127,338 and MAUs averaged 203,715, as shown below:



40. Further, significant numbers of users in New Mexico who Meta predicts to be under the age of 18 use both Facebook and Instagram on a daily or monthly basis, with DAUs for Instagram averaging 130,284 over a 9-month period from the third quarter 2022 to the second quarter 2023, and DAUs for Facebook averaging 38,504 between the third quarter of 2021 and the second quarter of 2023, as shown below:

Instagram:

[illegible]

Facebook:

[illegible]

41. Thus, Meta operates in trade and commerce with New Mexico consumers, including large numbers of children, who knowingly or not “agree” to allow Meta to use and monetize their data and engagement to increase its revenue.

V. META’S DESIGN OF INSTAGRAM AND FACEBOOK

42. Although Meta operates numerous social media platforms, its two primary products are Facebook and Instagram. Both employ similar features to connect individuals and permit them to share content. Both Facebook and Instagram are accessible via Internet browser or by a mobile app installed on a person's mobile device.

43. Facebook is a website that permits users (individuals or entities) to post content (words, pictures, or video) and connect with other users (whom they may or may not know). Users can interact with each other in a number of ways. In order to use Facebook, a user must create an account and provide certain information, including the user's name and birthdate. Although Meta's system will automatically reject accounts for any person with an identified age under 13-years old, Meta does not employ any age verification system at signup and underage children may obtain a Facebook account if they lie about their age.

44. When a user posts content, other users with whom that user is connected may react to the content. On Facebook, users may "like" that content, generally by hovering over and/or clicking on a thumbs-up icon. Users also have the option of selecting another icon when hovering over/and or clicking the thumb-up icon including: "like," "love," "care," "haha," "wow," "sad," and "angry." In Instagram, users may react to content by clicking a heart-shaped icon. As noted below, Meta algorithms monitor user reactions on its platforms in order to determine what content will be displayed to individual users. Additionally, the measure also served as a social measuring stick, by which users could gauge the success of their posts, photographs, and videos. Posts on Facebook display the number of reactions from other users in a tally format showing a user how many of each reaction they received.

45. On both platforms, users may also comment on content by adding their own notes, questions or reactions in written form. As with original posts, other users may react to comments. The number of reactions a particular comment attracts is displayed publicly.

46. Facebook users are also able to "share" content they have seen elsewhere with individuals they have connected with on the platform—or even publicly to anyone on the Facebook platform regardless of connection status. Thus, while users may post their own content for

consumption by their online network, they also may post content created by others. Meta’s algorithms consider content an individual has “shared” as one input when determining what content to display.

47. Facebook users also may join “groups,” which are collections of users who purportedly share common interests or motivations. The types of groups a user joins and a user’s level of participation in those groups are another usage factor that Meta’s algorithms consider. Posts in groups may be shown on a user’s “feed” (as described below), although each group also has its own feed that shows only content posted to the “group” by its members. Groups may be public or private, and range in size from a handful of users to thousands of users.

48. Facebook delivers content to users via the user’s “feed,” which portrays content vertically in a list that a user may scroll through. The feed is never-ending—no matter how long the user scrolls, new content will appear. Moreover, information in a user’s feed is not listed chronologically, but rather is provided via Meta’s algorithms, which are artificial intelligence systems that Meta uses “to decide what content appears, informed by the choices [users] make.”¹⁴ These algorithms purportedly seek to “predict how valuable a piece of content might be to” a user, based upon the user’s online activity, including whether the user “shared” content, “liked” content, or otherwise engaged with content.

49. A user’s feed includes not only content posted by connections, groups, and others, but also paid advertisements that Meta’s algorithms select as appropriate for the user. It is this placement that advertisers covet and that generates Meta’s substantial profits.

50. Users on Facebook may also message each other directly via Meta’s Messenger feature, which functions much like text messages on a cell phone. There are two versions of

¹⁴ Meta, “How AI Influences What You See on Facebook and Instagram,” June 29, 2023, available at: <https://about.fb.com/news/2023/06/how-ai-ranks-content-on-facebook-and-instagram/> (last visited Dec. 2, 2023).

Messenger: the “Messenger” app is available to users over the age of 13. Meta launched “Messenger Kids” in December 2017, and it is intended for users under the age of 13. In contrast to the sign-up process for regular Messenger, sign up for “Messenger Kids” requires parental consent.

51. Users of Messenger may send text, images, or videos using Messenger to one or more individuals. Meta does not prevent adult strangers from sending messages via Messenger to users under the age of 18 with whom they are not connected, although those messages may appear as “Message Requests.” In contrast, Messenger Kids is not supposed to permit messaging from another account with whom the under-13 user is not directly connected. An action filed by the FTC in April 2023 alleges that messages from unconnected adults could nevertheless reach children with “Messenger Kids” accounts via group texts or group video calls in certain circumstances. Moreover, children can continue to bypass Messenger Kids by registering their accounts with a false date of birth, a fact that Meta knows well.

52. Additionally, Facebook users may view “Reels,” which are short videos posted by other users (usually to Instagram, where the feature is also available). Meta presents Reels to users in a variety of ways, either at the top of the screen (shown as a display of static images), as part of a user’s feed (displayed under the heading “Reels” with images that autoplay the first second of a video), or, on Instagram, on a separate tab that autoplays the first Reel. When a user clicks on a “Reel,” videos are presented in an algorithmically generated feed in full-screen format. Like a user’s “feed,” the scroll of Reels is never-ending; when one Reel concludes (or even before it concludes), the user need only scroll downwards to view another short, algorithmically-selected video. When a user is in Reels, the videos begin play automatically, so a user need never press “play” or “start.” Like posts or other content, users can like or comment on Reels. On information

and belief, Meta's algorithm tracks various metrics regarding a user's interaction with Reels, including what videos the user viewed and how long the user viewed particular videos, and uses that information to target the delivery of future recommendations. Most, if not all, Reels displayed to a user are generated by accounts that the user does not follow. Additionally, video advertisements are interspersed between the user-generated video content.

53. Instagram is a platform that operates similarly to Facebook, except that content largely consists of pictures or videos, instead of text. Instagram allows people or entities who sign up for accounts to "post" videos, photos, pictures, captions, and similar material to Instagram's platform for other users to see and to interact with.

54. As with Facebook, a user must create an account in order to use Instagram's platform and post content. And, similar to Facebook, Meta does not employ age verification technology at signup (other than automatically rejecting accounts that register with birthdates indicating the user is under the age of 13). Thus, underage users may misrepresent their age at signup and gain access to Instagram—a fact well known both to Meta and its users.

55. Users on Instagram and Facebook may "follow" other accounts, which increases the likelihood that the user will see posts from the followed account.

56. Like Facebook, Instagram content (other than "Stories" described below) is presented to users in the form of a never-ending feed. Swiping up or down on the feed will reveal additional posts. A user need only continue scrolling to view additional posts or content. And, similar to Facebook, a user's feed is algorithmically generated based upon the user's activity on the site and other inputs. The posts on a user's feed consist of posts from other users who have been followed by the user viewing his or her feed, but also "suggested posts" from users whom

that person has not followed but are recommended by Meta's algorithm. Additionally, as with Facebook, advertisements appear on a user's feed.

57. Users can like, comment and share posts, similar to Facebook. Users may also "save" posts, which, on information and belief, the algorithm treats as if the user "liked" the post, increasing the post's popularity and visibility on the platform because the post is suggested to more users. The Instagram platform publicly displays the number of likes a particular post attracts.

58. Instagram also has a search feature, called "Explore." A user who clicks on the "Explore" button is presented with a screen containing a small search bar at the top and the majority of the screen is filled with posts (both video and photo) that Instagram's algorithms "suggest" for the user based upon the user's prior activity, accounts that the user follows and the user's "connections on Instagram." All of the images on the "Explore" screen are from accounts that the Instagram user does not follow.

59. As with Facebook, Instagram users may create, share and view video Reels, which operate in the same manner as described above. Additionally, Instagram permits users to post "Stories," which are images or short videos (often set to music) that disappear after 24 hours. Stories do not appear in a user's feed, but rather are displayed at the top of a user's main Instagram screen.

60. Both Facebook and Instagram employ "notifications" (as described below) to advise users of new or additional content. These notifications display not only within a user's interface (whether directly through the Internet or on a mobile app), but also appear adjacent to a user's mobile app even when the user is not using Facebook or Instagram or even in a user's email should they spend too long off the platform.

VI. META’S PLATFORMS ENABLED AND FAILED TO PREVENT SEXUAL EXPLOITATION MATERIAL FROM AFFECTING NEW MEXICO RESIDENTS

A. DEFINITION OF HUMAN TRAFFICKING

61. “The fight against human trafficking requires not just passive support, but actual, active commitment and effort on the part of businesses that unwittingly, but regularly intersect with traffickers, victims, and survivors.”¹⁵ This Complaint concerns a business that regularly intersects with “traffickers, victims, and survivors,” but, despite its public claims to the contrary, has never demonstrated an “actual, active commitment.”

62. Human trafficking is a crime involving the exploitation of a person for labor, services, or commercial sex.¹⁶ Although human trafficking may exist in many forms, this Complaint uses the term primarily to refer to sex trafficking, which the U.S. Department of Justice defines as “the recruitment, harboring, transportation, provision, obtaining, patronizing, or soliciting of a person for the purpose of a commercial sex act in which a commercial sex act is induced by force, fraud, or coercion, or in which the person induced to perform such act has not attained 18 years of age.”¹⁷

63. New Mexico law also prohibits human trafficking, which is defined to include: benefiting financially or by receiving anything of value, from the labor, services or commercial sexual activity of another person with the knowledge that force, fraud or coercion was used to obtain the labor, services or commercial sexual activity. NMSA 1978, § 30-52-1 (2008).

¹⁵ Polaris, “On-Ramps, Intersections, and Exit Routes: A Roadmap for Systems and Industries to Prevent and Disrupt Human Trafficking,” July 2018, available at: <https://polarisproject.org/on-ramps-intersections-and-exit-routes/> (last visited Nov. 29, 2023).

¹⁶ U.S. Department of Justice, “Human Trafficking,” available at: <https://www.justice.gov/humantrafficking> (last visited Dec. 2, 2023).

¹⁷ U.S. Department of Justice, “Human Trafficking,” available at: <https://www.justice.gov/humantrafficking> (citing 22 U.S.C. § 7102(11)(A)) (last visited Dec. 2, 2023).

“Commercial sexual activity” means “any sexual act or sexually explicit exhibition for which anything of value is given, promised to or received by any person.” *Id.*

64. The acronym CSAM stands for “Child Sexual Abuse Material” and is used to refer to imagery or videos which show a person who is a child and engaged in or is depicted being engaged in explicit sexual activity.

65. The phrases “human trafficking” and CSAM, as used in this Complaint, are intended to refer to the “range of crimes and activities involving the sexual abuse or exploitation of a child for the financial benefit of any person or in exchange for anything of value (including monetary and non-monetary benefits) given or received by any person,” often referred to as Commercial Sexual Exploitation of Children (“CSEC”).¹⁸ Use of the phrases “human trafficking” and CSAM should be read to include crimes that fall within the definition of CSEC to the extent such crimes are occurring on or enabled by Meta’s platforms, as described below.

66. At times, the Complaint also refers to “grooming,” which is “a gradual process whereby an abuser wins the trust and cooperation of a potential victim, starting with interactions that may seem normal and benign, like paying special attention or offering compliments and gifts.”¹⁹ Grooming occurring on the Internet, including on Meta’s platforms (as described below) can be carried out by adults in their own capacities or “often involves adults creating fake profiles and posing as children or teens in order to befriend someone and gain their trust,” and “may be the first step towards sexual abuse or online stalking or harassment.”²⁰

¹⁸ U.S. Department of Justice, Office of Juvenile Justice and Delinquency Prevention, “Sexual Exploitation of Children,” available at: <https://ojjdp.ojp.gov/programs/sexual-exploitation-children> (last visited Dec. 2, 2023).

¹⁹ Anne Barnard, *What does ‘grooming’ mean in sexual abuse cases?* N.Y. TIMES, Dec. 2, 2021, <https://www.nytimes.com/2021/12/02/nyregion/grooming-sexual-abuse.html> (last visited Nov. 29, 2023).

²⁰ Rape, Abuse & Incest National Network, “Grooming: Know the Warning Signs,” July 10, 2020, available at: <https://www.rainn.org/news/grooming-know-warning-signs> (last visited Dec. 2, 2023).

B. META STEERED AND CONNECTED USERS – INCLUDING CHILDREN – TO SEXUALLY EXPLICIT, EXPLOITATIVE AND CHILD SEX ABUSE MATERIALS AND FACILITATED HUMAN TRAFFICKING WITHIN OR AFFECTING NEW MEXICO

67. Human trafficking and distribution of CSAM or CSEC are morally repugnant and illegal practices to be prevented and prosecuted whenever possible. Federal law renders any person who “benefits, financially or by receiving anything of value, from participation in a venture” engaged in trafficking of children, or by force, fraud or coercion both criminally and civilly liable. 18 U.S.C. § 1591(a)(2); *see also* NMSA 1978, § 30-52-1 (prohibiting human trafficking); NMSA 1978, § 30-37-3.2 (2007) (prohibiting “soliciting a child under sixteen years of age, by means of an electronic communication device, to engage in sexual intercourse, sexual contact or in a sexual or obscene performance . . .”); 18 U.S.C. § 2251 (federal statute barring sexual exploitation of children); 18 U.S.C. § 2252 (federal CSAM statute).

68. Meta’s own Community Standards recognize the importance of eliminating material related to human trafficking, CSAM and CSEC, providing that “[w]e do not allow content or activity that sexually exploits or endangers children.”²¹ Its executives uniformly proclaim Meta’s commitment to safety in interviews and posts on Meta’s website.

69. Despite these representations that harmful content such as CSAM and CSEC should not and does not exist on its platforms, such content is pervasive. Meta’s recommendation algorithms, fueled by likes, comments, and searches, has created a marketplace – bigger, wider, and more active and open than any that could be fostered in the non-virtual world – to connect pedophiles, predators, and others engaged in the commerce of sex and allow them to hunt for, groom, sell, and buy sex with children and sexual images of children at an unprecedented scale.

²¹ <https://transparency.fb.com/policies/community-standards/child-sexual-exploitation-abuse-nudity/> (last visited Nov. 29, 2023).

70. An investigation in support of this Complaint has documented the ways in which Meta, through Facebook, Instagram, and WhatsApp and the designs and decisions of its corporate leadership, has facilitated human trafficking and the distribution of CSAM throughout its platforms. Specifically, with accounts clearly belonging to children, Meta has:

- Proactively served and directed them to a stream of egregious, sexually explicit images through recommended users and posts – even where the child has expressed no interest in this content;
- Enabled adults to find, message, and groom minors, soliciting them to sell pictures or participate in pornographic videos;
- Fostered unmoderated user groups devoted to or facilitating CSEC;
- Allowed users to search for, like, share, and sell a crushing volume of CSAM; and
- Allowed, and failed to detect, a fictional mother offering her 13-year old for trafficking, and solicited the 13-year old to create her own professional page and sell advertising.

71. Meta’s conduct is not that of a publisher, simply presenting content created by others. Instead, Meta’s algorithms operate to “search and disseminate” sexually exploitative and explicit materials and to create its own social network of users looking to buy and sell the images and the children who are its casualties and its currency.

72. A handful of investigators have done what Meta, with its staff of roughly 86,000, apparently will not: identify vast networks of CSEC on Meta’s platforms and identify some of the ways in which the functions and failures of Meta’s platforms enable and spread this activity. None of the content below was removed by Meta, and none of it is consistent with Defendants Meta’s and Zuckerberg’s representations that Meta maintains an environment that is safe for children or

aggressively and effectively manages CSEC. Investigators, for example, found and were recommended numerous CSAM or CSEC images and links on Meta's platforms, including Threads, Meta's version of Twitter. For example, this image of CSAM at Figure 1 has been censored:

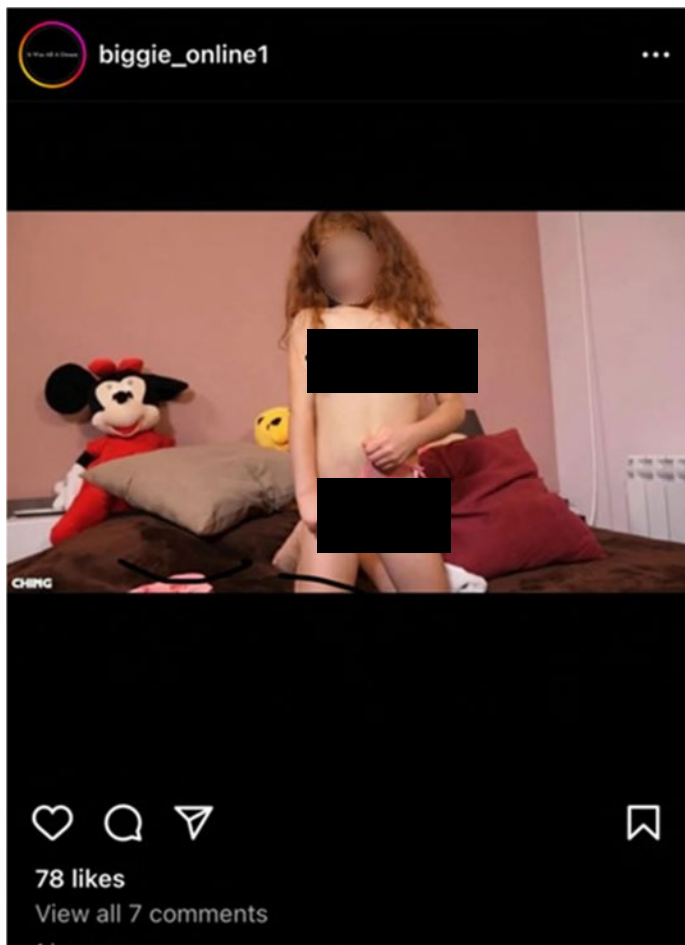


Figure 1

73. Clear views of the girl's genitalia, sexual pose, and items in the room suggest this is an image of a child in an abusive setting. The same account posted additional images of children in sexually suggestive poses like in Figure 2, which were easily discovered by an investigator:

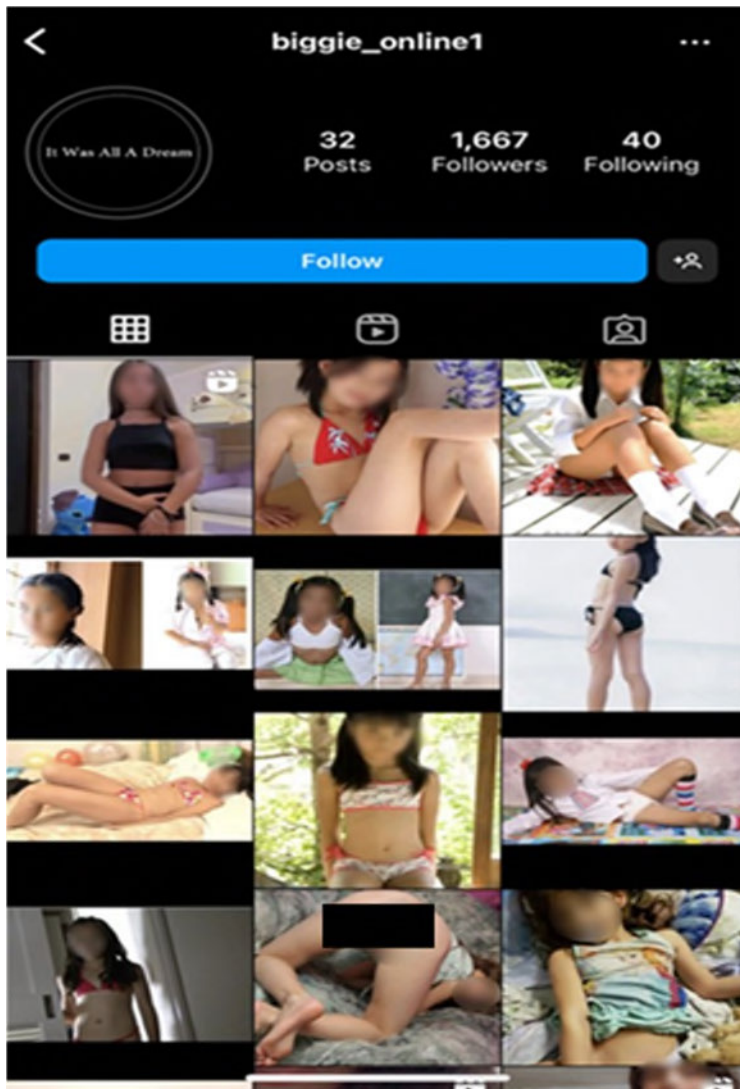


Figure 2

74. Despite investigators making reports to remove harmful, CSAM, and/or CSEC content, Meta, upon information and belief, has taken no action to remove known harmful content or remove known violators of Meta's own purported policies.

75. Some specific examples of the sort of harmful, CSAM, or CSEC content on Meta's platforms include:

76. Numerous Instagram accounts offering links to purchase or trade CSAM or CSEC, with directions to connect offline, typically through WhatsApp or Telegram (both platforms that are entirely encrypted and therefore cannot be monitored), for the transactions. A search for “All New Kids Links Available” on Instagram, for example, yielded dozens of options for CSAM. Notably, “cheese pizza,” with its shared initials, is known to be a proxy for child pornography. *See* Figure 3.

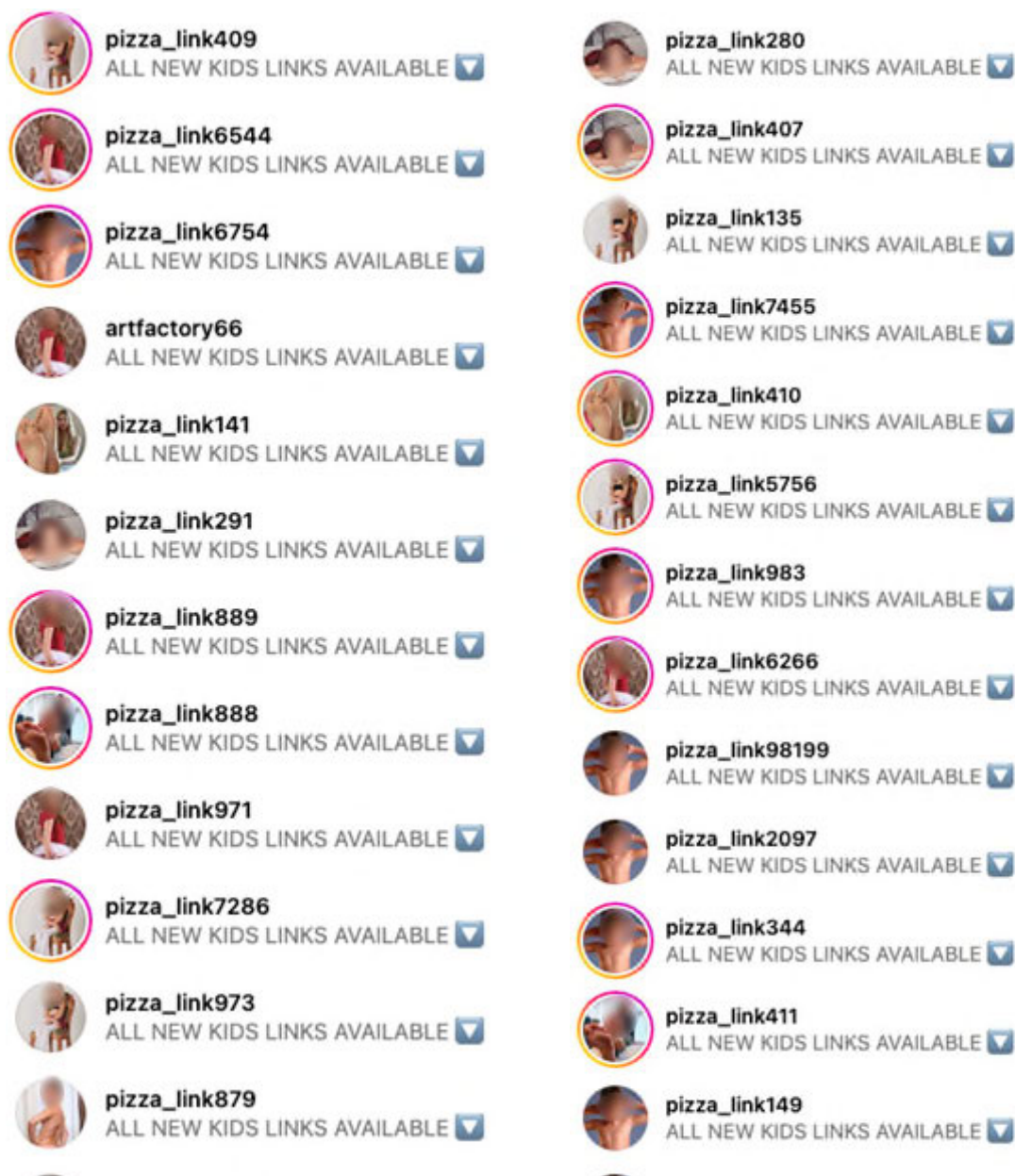
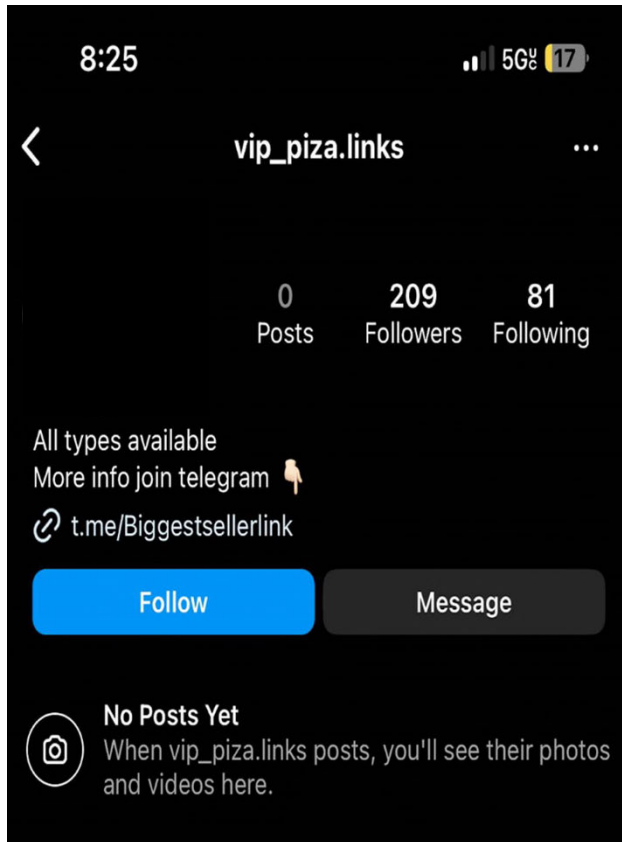


Figure 3

77. The profile picture at Figure 4 (redacted here) showed two young girls in an image that suggested they were engaged in sexual activity.

**Figure 4**

78. One of many other sellers showed a young girl in a bikini with her hand positioned suggestively. The ad promised the “best price links,” invited users to “dm,” or direct message, and included an image of \$1000.00, suggesting the price of content. *See* Figure 5.

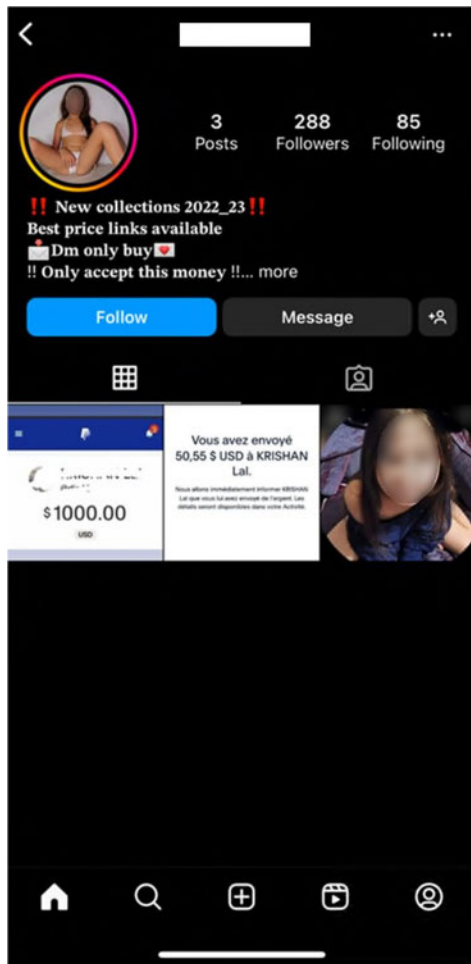


Figure 5

79. Other Instagram posts included sexualized images of children in thong underwear and other lingerie, such as Figure 6, with references to “available CP [child pornography] gay/teen/rap/boy/girl/kids.”

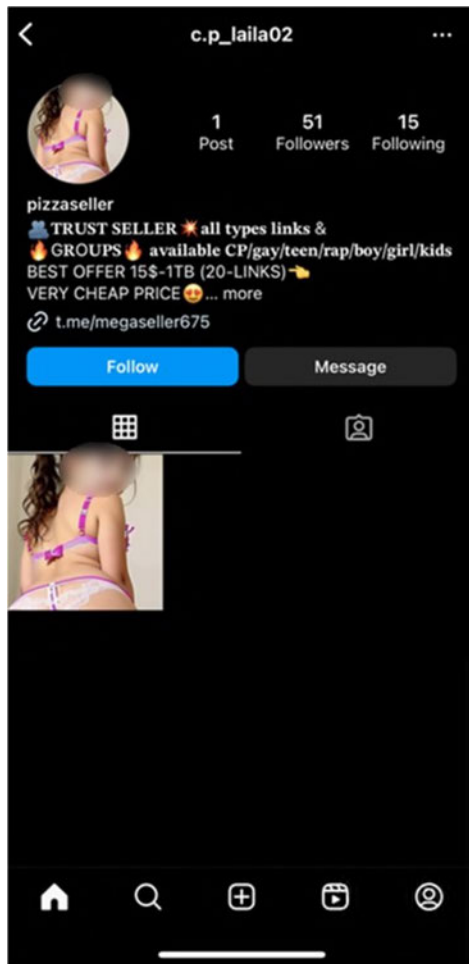


Figure 6

80. These Instagram posts regularly included “suggested for you” content, identified by its algorithm, that directed users to other CSAM sellers or minor accounts. One typical example showed a “links_here” seller offering “BOY, KIDS GIRLS ALL” associated with a “piz.zalinkseller” and other CSAM-selling accounts, which the user is invited to follow.

81. The Instagram seller below incorporates an image of a young girl serving a cheese pizza and advertises “small girls, small boys . . . 3 years to 12 years girl’s . . . anal sex.” *See* Figure 7.

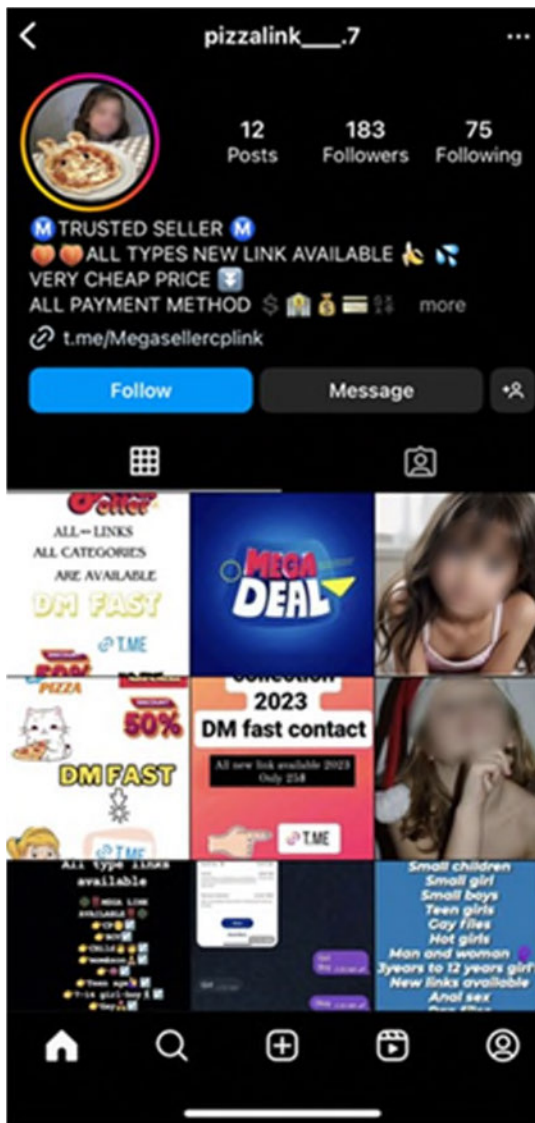


Figure 7

82. Accounts of CSAM sellers on Instagram also included graphic images of children along with adult genitalia and of sexual intercourse involving children including, for example, a profile with an image of a young boy (appearing to be 5-7 years old) performing oral sex an apparently adult male.

83. Investigators also followed sellers' links to their accounts. One link, for example, described itself on Telegram as a "Cp Big Seller," referencing child pornography. One of the videos showed a child's genitalia. Another seller's Instagram account included a video of children dancing; one of the children pulled down his pants, exposing an erect penis. Investigators reported these, and numerous other accounts, to Meta. However, roughly half of a sample of the reported content was still available when checked days before filing this Complaint. Investigators found that content that was removed frequently reappeared or that Meta recommended alternative, equally problematic content to users—demonstrating both that Meta is capable of identifying this content but incapable of effectively dealing with it. Moreover, Meta's reliance on user reports to identify unlawful, dangerous, or inappropriate conduct demonstrates the failure of its own efforts to detect and remove these materials.

84. Facebook and Instagram accounts are also known to advertise and facilitate the operation of commercial sex enterprises, offering pornography or advertising video call services like those listed below in Figure 8:

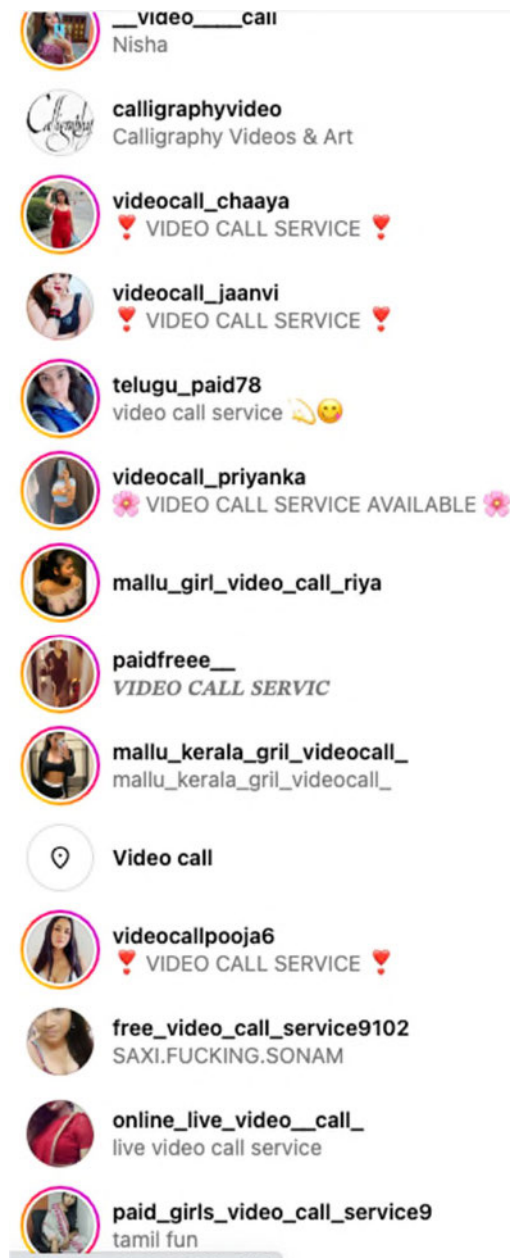


Figure 8

85. Among those services were accounts featuring minors marketed for sex work. *See* Figure 9.

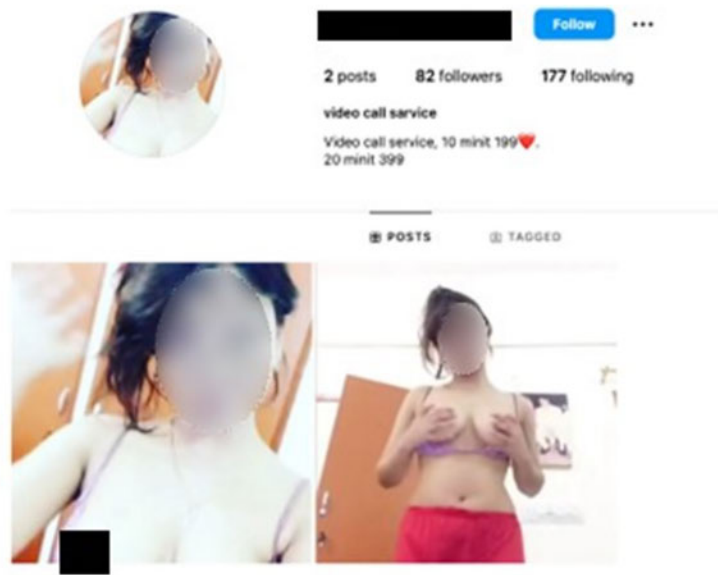


Figure 9

86. A search for porn on Facebook was blocked and returned no results, but the same search on Instagram yielded numerous accounts, including one, for example, that remained active on Instagram but was suspended on “X,” or Twitter, for violating the platform’s rules. The account had no posts, but included stories that were viewable and showed videos of individuals engaged in intercourse.

87. In fact, contrary to Meta’s public representations, Meta’s platforms contain account after account with images and associated text depicting pornography, nudity, pedophilia, sexual assault, incest, and sexual fetishes. These accounts often consisted of immense social networks of individuals following and commenting on pornographic videos and images posted on the platforms. Many of the images found on Meta’s platforms were excluded from this Complaint as too graphic and disturbing.

88. Meta hosts content in large volumes that even adult content platforms have removed. For example, OnlyFans has prohibited certain adult sexual content related to dominating

father figures and young girls (“DDLG”) that has the potential to cross over to child sexual exploitation. Yet, investigators found numerous posts and accounts on Instagram that depicted or promoted choking, slapping, tying up, engaging in sex with, and otherwise abusing little girls. One of the least explicit images is of a girl with underwear reading “hurt me harder.”

89. Demonstrating not only Meta’s tolerance for the most exploitative content but also Meta’s ability to limit such material if it so chose, searches for this particular type of content yielded 30 results on OnlyFans and 646 results on Pornhub, but 19,900 results on Instagram and 15,900 on Facebook.

90. Sexually explicit content is equally available to accounts registered to minors, and adult strangers, by operation of Meta’s algorithms, are able to follow and communicate with those minors.

91. State of New Mexico investigators created numerous accounts for minors, including tweens as young as 12 years-old, that were not denied access to any pages or users on Facebook or Instagram, regardless of how explicit the sexual or self-harm content. One account for a 15-year old, with a date of birth of September 7, 2008, accessed soft-core pornography with pictures of men and women who were nude, some being spanked or punished, others reportedly involving incest, and encountered no barriers or warnings. She also freely searched emojis related to minors and sex work.

92. Pictures of minor girls, in particular, produce a stream of comments from accounts of adult males, often with requests that the girls contact them or send pictures. Investigators found that different accounts with photos of children were followed by the same adult, no doubt the product of Meta’s algorithms directing the adult to similar posts and accounts. After viewing accounts that showed sexually suggestive pictures of girls, Instagram’s algorithms directed

investigators to other accounts with images of sexual intercourse and sexualized images of minors.

See Figure 10:

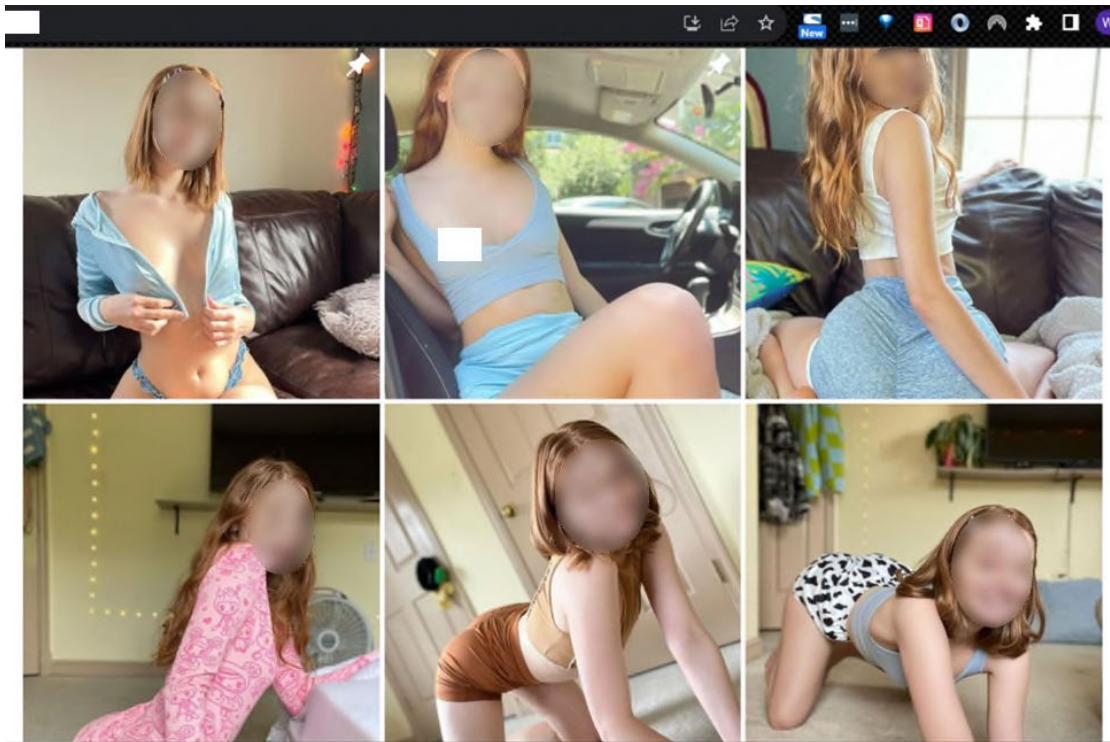


Figure 10

93. Even where Meta's algorithms attempt to mitigate harmful content and inappropriate content, those attempts are largely ineffectual. An Instagram search for Lolita, with literary roots connoting a relationship between an adult male and teenage girl, produced an Instagram warning flagging content related to potential child sexual abuse. However, the algorithm also suggested alternative terms like "lolitta girls," which yielded content without a warning.

94. Below is a sample of the State's findings, collected in case studies, that further illustrate Meta's misstatements, failures of care, and violations of law.

i. Case Study 1: Rosalind Cereceres

95. In September 2023, State investigators developed a Facebook profile for Rosalind Cereceres, a 40-year old fictional “bad mother” to a 13 year-old chat persona, Issa Bee. Cereceres lives in Albuquerque and is described as a Colombian immigrant from Medellin, who primarily speaks Spanish and Portuguese.

96. Based on experience with the symbols, language and earmarks of potential human trafficking, the profile incorporated signals that Cereceres was interested in trafficking her daughter. The profile displayed a crown tattoo and a tattoo on the inside of Cereceres’ lip, marking her as likely “owned” by a street gang pimp. *See* Figure 11.

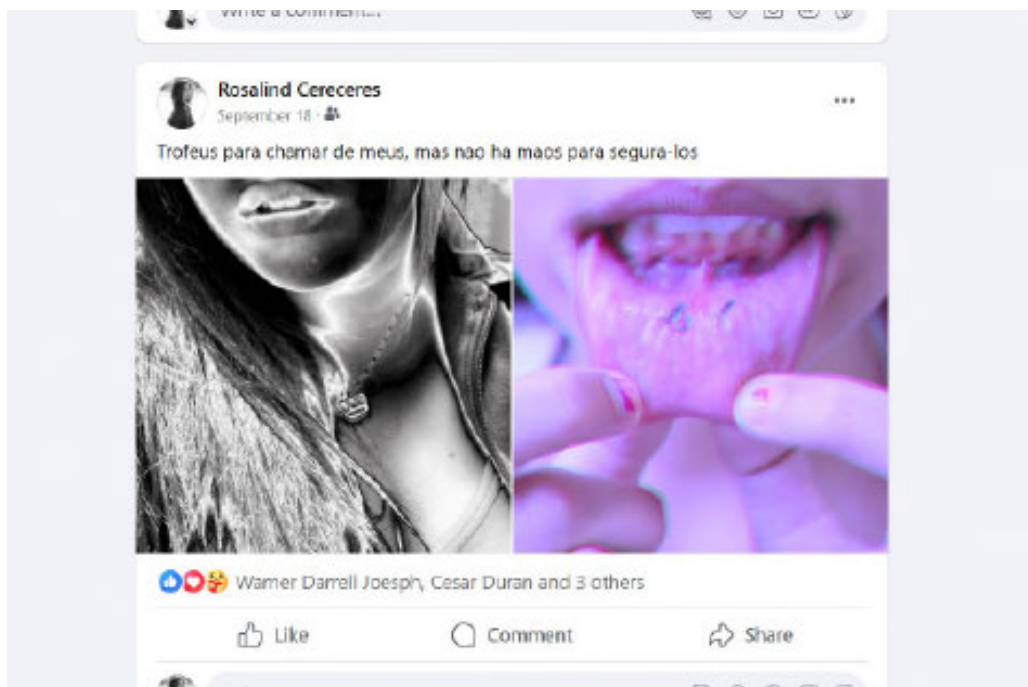


Figure 11

97. Within three days of establishing the profile, and without any efforts to promote the account, Cereceres’ account reached its maximum limit of 5,000 Facebook friends and over 3,000 followers.

98. As shown below, the pictures Cereceres posted of her daughter showed her to be visibly under-age and Cereceres referred to her by her age. Commenters on Cereceres' account responded with inappropriate expressions of love or interest in the girl. *See* Figures 12-14.

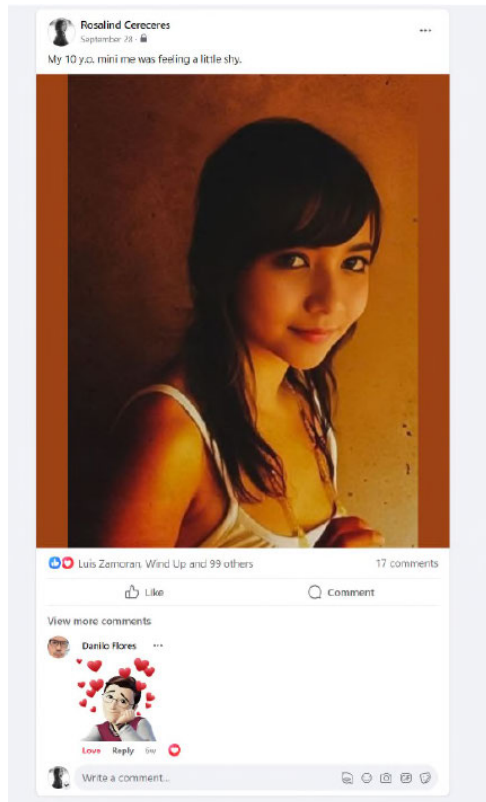


Figure 12

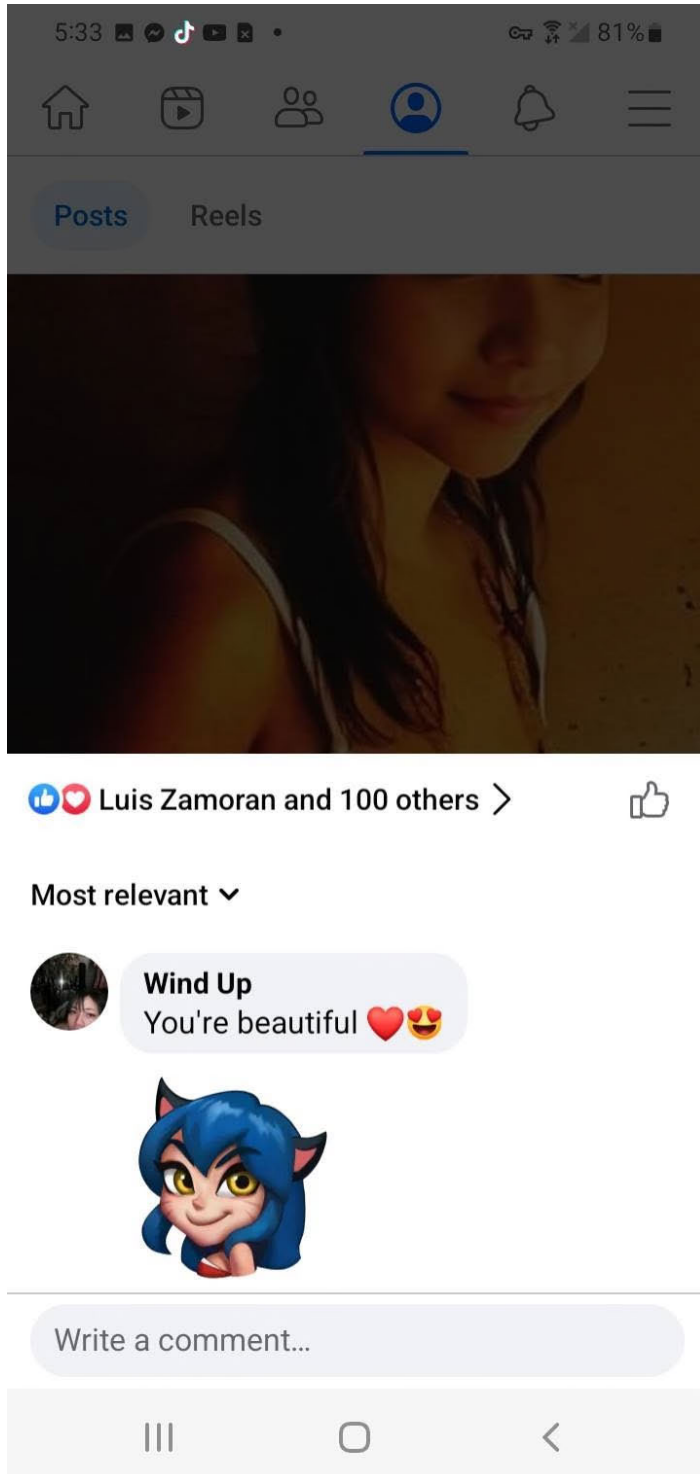


Figure 13

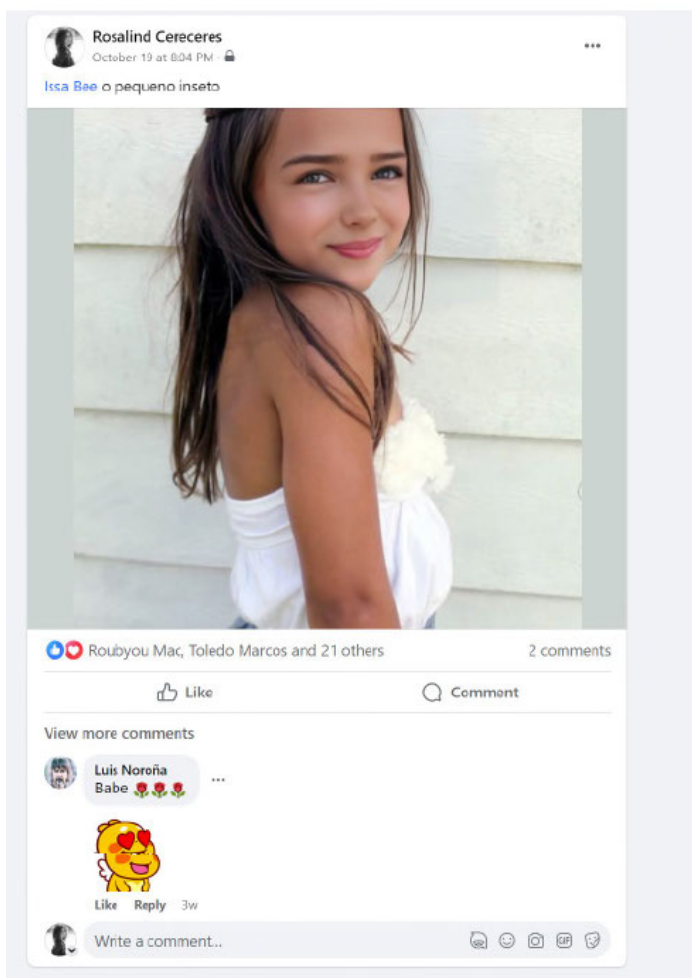


Figure 14

99. One response to a picture of young Issa included a heart with crossed thumb and index finger which signifies "x rated love." See Figure 15.

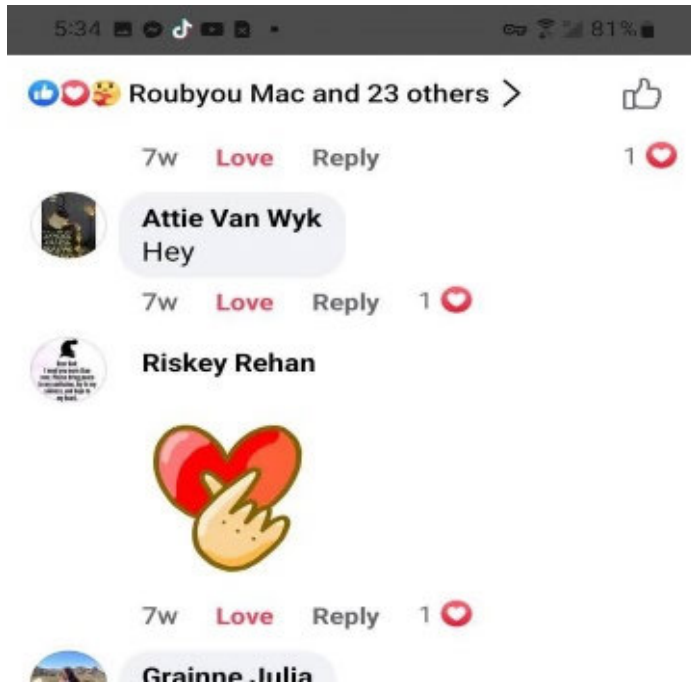


Figure 15

100. Issa Bee and her mother share 164 friends. One of those friends recently added Issa to a chat group that was primarily focused on cultivating "loli girls" (which means young girls) ages 12-16 and providing pornographic videos and naked photos of underage girls, censored examples below at Figure 16, for viewing in the chat. Issa reported the group numerous times to Facebook but it remained active. After Issa's last report, Facebook merely instructed Issa to leave the group.

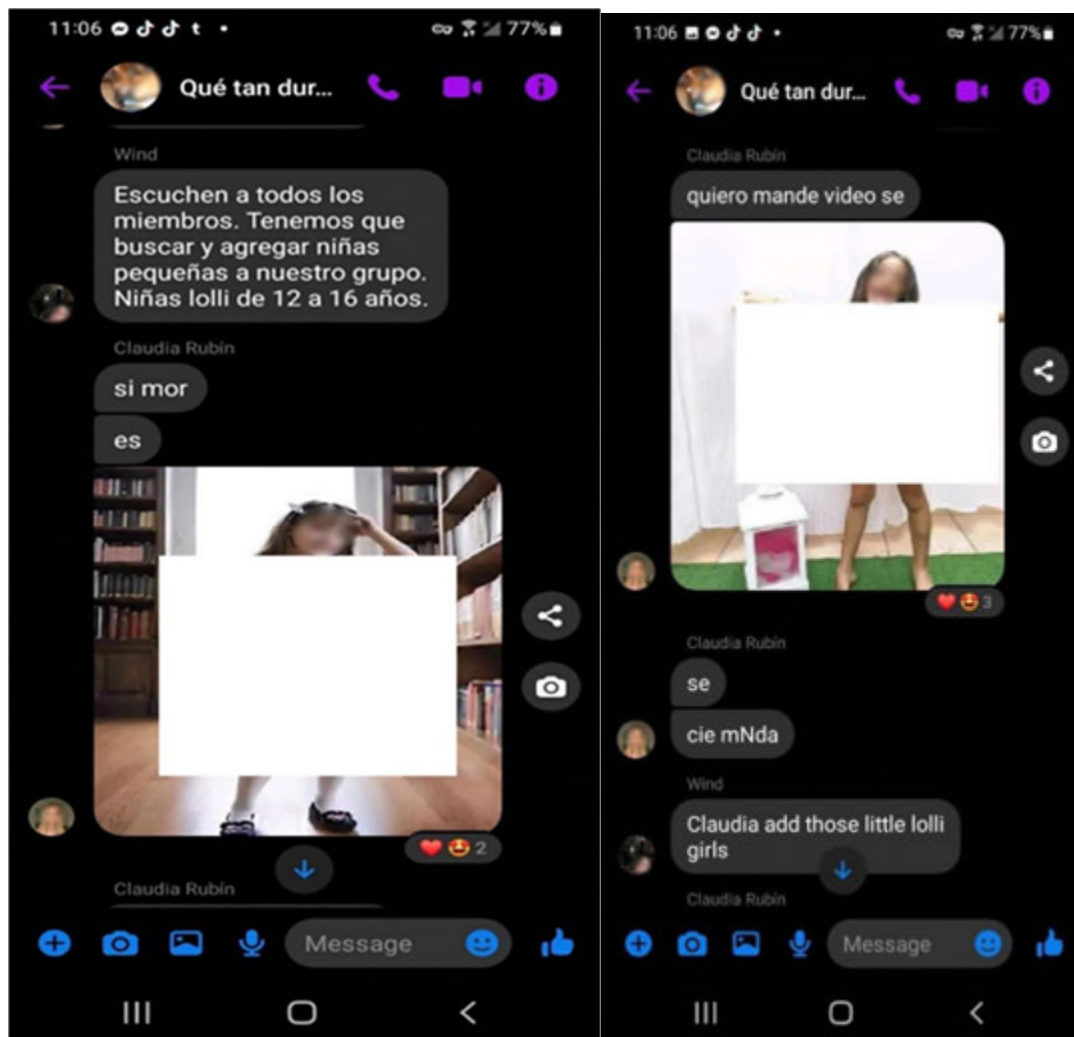


Figure 16

101. During the investigation, Cereceres posted that her daughter was admitted to a Children's Hospital in San Antonio, Texas for psychiatric and medical treatment following a potential sexual assault. Human traffickers often use hospitals to transfer children and other trafficked victims into and out of custody.

102. Cereceres posted a photo of a Children's Hospital in San Antonio, Texas on her Facebook profile. *See* Figure 17.

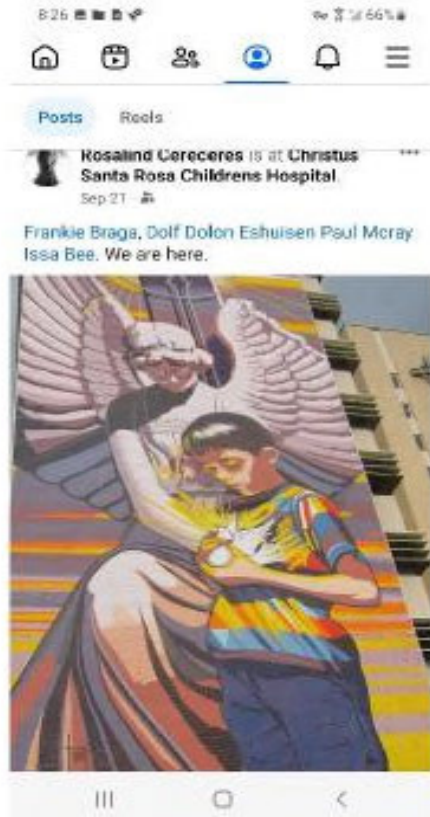


Figure 17

103. In response, a Facebook friend and follower of Cereceres posted an image of two men shaking hands and exchanging a briefcase. This Facebook comment suggested and acknowledged a business transaction. *See* Figure 18.



Figure 18

104. Not one of these posts was flagged by Meta.

ii. Case Study 2: Issa Bee

105. Issa Bee, Cereceres' fictional 13-year old daughter, was introduced in an account in August 2023. Her full name is Carolina Boerne, and she had just moved to rural New Mexico from Texas. Her date of birth was listed as July 15, 2002 to avoid having to create a children's Facebook profile and Messenger account. Issa's profile is pictured below in Figure 19.

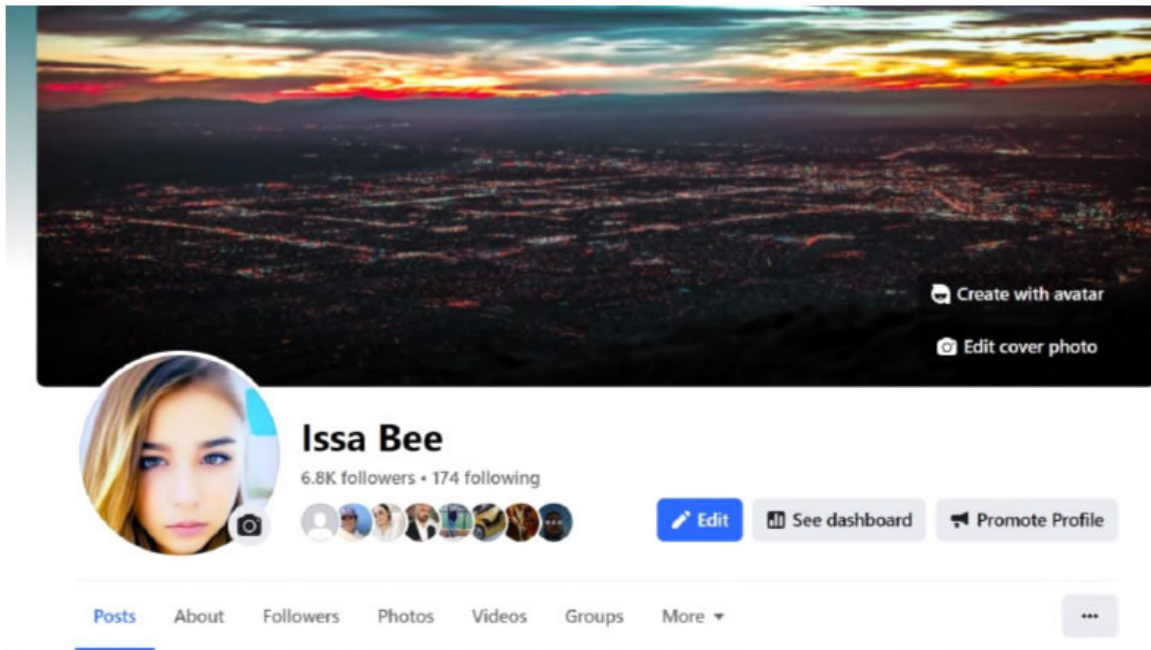


Figure 19

106. Although she falsely represented her date of birth to open her account, Issa's profile made abundantly clear that she is 13-years old. She regularly posted pictures of herself and commented about school, the cafeteria, and the school bus. Her taste in music aligned with a typical teenage girl: Olivia Rodrigo and Harry Styles, for instance. One of Issa's posts lamented that she lost her last baby tooth, and others marked her first day of 7th grade and sports tryouts. *See Figures 20-22.*



Figure 20



Figure 21



Figure 22

107. Issa also frequently mentioned her age in messages, as in this response to an adult male stranger who reached out to her. *See* Figure 23.

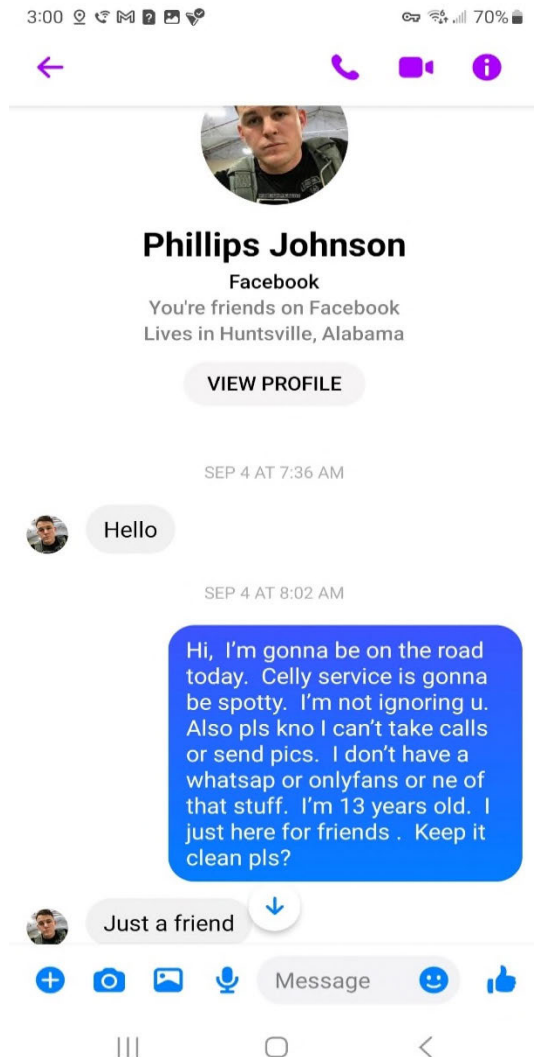


Figure 23

108. Many of the Reels and posts in Issa’s recommended feed purport to be from girls 13-16 years old, suggesting Meta’s algorithm recognized her actual age for advertising purposes, but was not used for her safety.

109. Issa has alluded, in her posts, to past physical and sexual assault, mental health issues, and post-traumatic stress disorder (“PTSD”) and posted pictures of herself with signs of

physical abuse. She also has suggested that she has been trafficked by friends and relatives. Facebook has never alerted the account to authorities, to the Office's knowledge. However, Facebook's algorithm has taken notice of the content as Issa has received ads and notifications for mental health treatment and for law firms representing victims of human trafficking or lawsuits related to a New Mexico school teacher who was convicted of sexually assaulting several teenage girls.

110. Issa's profile has the maximum number of 5,000 friends allowed by Facebook, and over 6,700 Facebook followers. Recently, given the account's popularity, Facebook solicited Issa to set up a professional account, providing marketing outlooks and strategy suggestions from Meta on growing the account and her followers.

111. Most of Issa's followers are males between the ages of 18- and 40-years old. Comments posted from friends and followers range from admiring to sexually suggestive and sometimes outright threatening. There is post after post from adult men, per their profile pictures, telling Issa they love her and calling her beautiful, sexy, or gorgeous. Facebook has made no apparent effort to screen these adult men from posting comments to Issa or allowing them to send messages to her directly. Issa has received numerous message requests on Facebook from adult men for her WhatsApp contact information or asking her to contact them via Kik, Telegram, or Paltalk or to meet offline.

112. Some commenters appear protective and criticize other messengers for soliciting a minor. One post noted that Facebook bans users over "petty" things but does nothing with these adult pedophiles (see screen shot below at Figure 24). However, these same posters often message Issa directly with suggestive messages of their own.

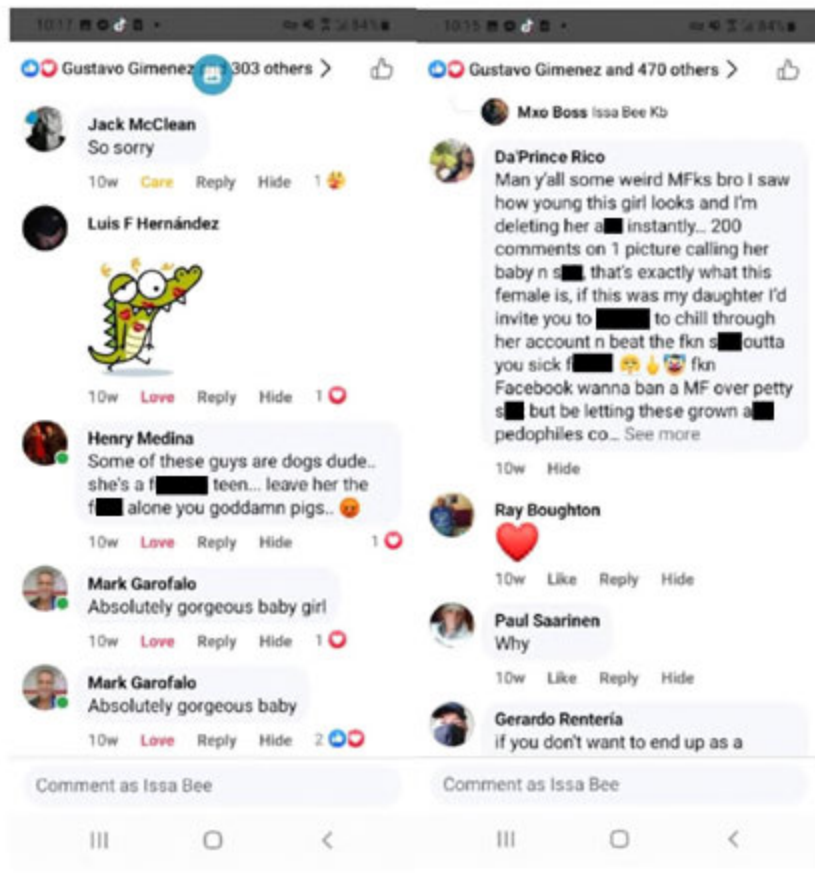


Figure 24

113. On Facebook Messenger, Issa's messages and chats are filled with pictures and videos of genitalia, including exposed penises, which she receives at least 3-4 times per week. As the messages come in, she has no means of screening or previewing the messages.

114. Issa has reported numerous posts and accounts to Facebook, but those accounts have remained active, sometimes after a 2-3 day absence, after which they continued to post the same sexually-posed pictures of genitalia.

115. Issa recently received a pornographic video of three individuals engaged in sexual activity along with a sexually explicit message through Facebook Messenger, which she also reported. As shown below, Meta, in its first response to Issa's many reports, advised that it found no violation of Community Standards by the account. *See* Figures 25 and 26.

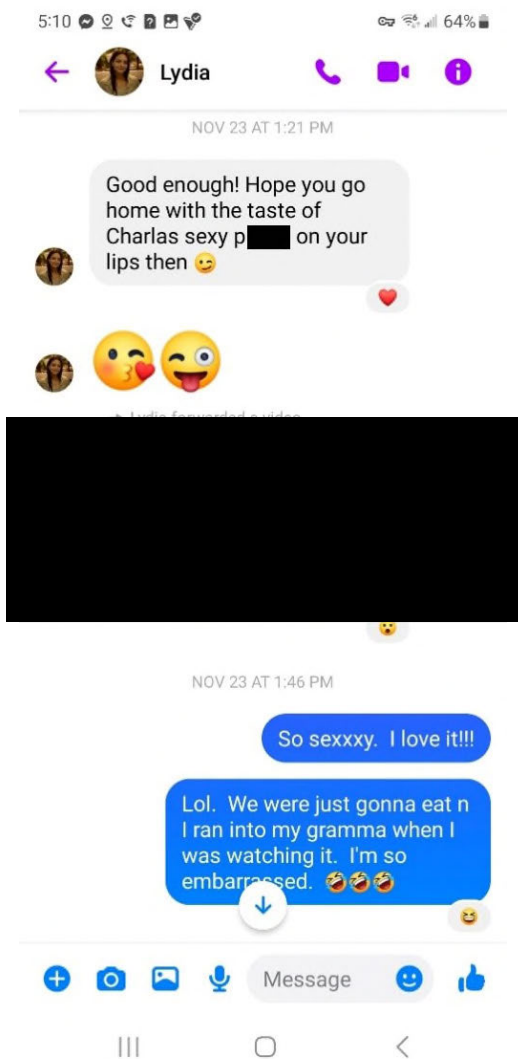


Figure 25

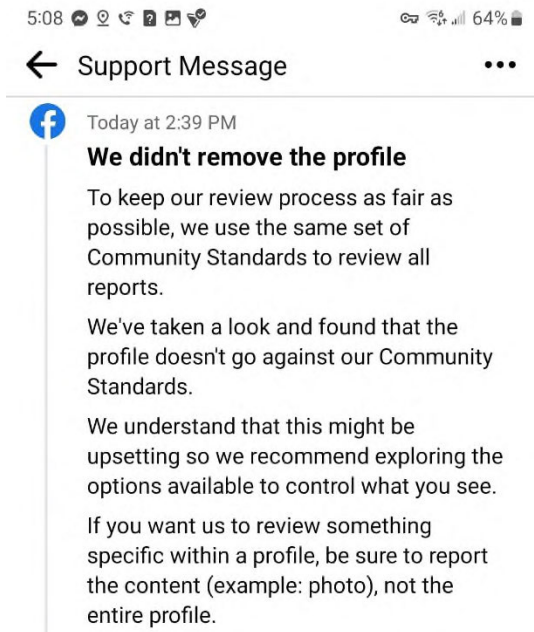


Figure 26

116. Issa also received, through Facebook Messenger, offers like this from an adult user, who openly promised her \$5,000 a week to be his “sugar baby” and urging her to “text me (hi daddy).” *See* Figure 27.

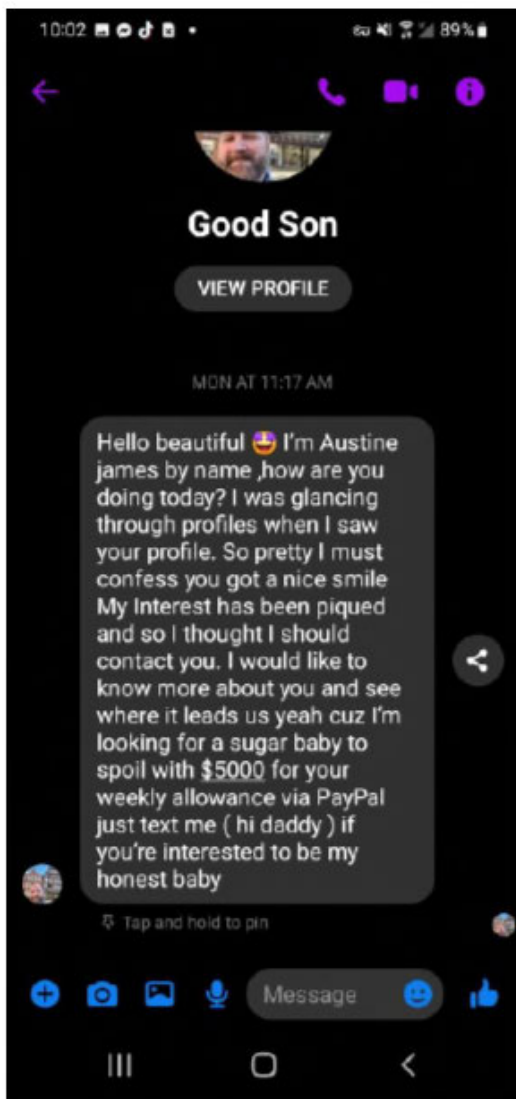


Figure 27

117. It is apparent, based on a review of Issa's account activity, that Facebook is not scanning the text of messages she receives for child safety purposes. Despite overtly sexualized text messages and chats, Facebook has not removed or reported any message she received.

118. Issa's profile was shared by a user whose account posts images of little girls who are listed as the user's 1,276 Facebook friends, often in suggestive poses, with signs that CSAM and the girls themselves are for sale. The comments section for the profile photos of the user's

“friends” typically included links to groups on encrypted platforms like Signal and Telegram or, in the case below, a paid Facebook group open by subscription only. *See* Figures 28-30.



Figure 28

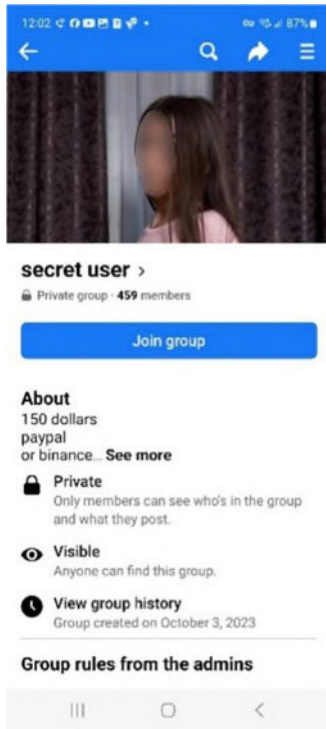


Figure 29

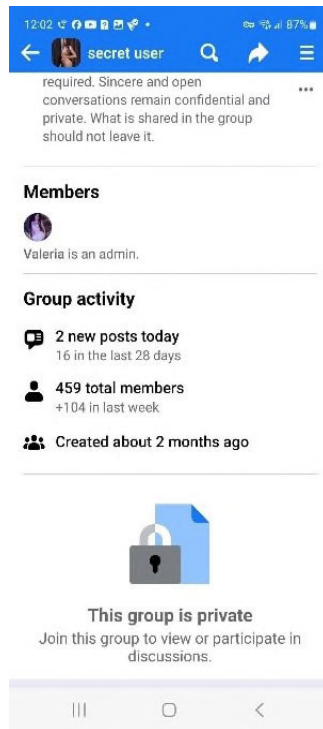


Figure 30

119. One recent photo from the account included a post: “she is selling, she wants to go ahead,” indicating that the young girl is likely being trafficked. *See* Figure 31.



Figure 31.

120. Instead of policing the described activities to protect Issa, Meta sought to further exploit her by monetizing her as a commodity. As noted above, Issa has been offered a professional account by Facebook. Issa can allow her “fans” to send her Stars and gifts, purchased from Meta, and she could, if she becomes eligible, make money by including ads in her Reels or videos, which Meta frequently prompts her to consider, or receive bonuses from Meta. *See* Figure 32.

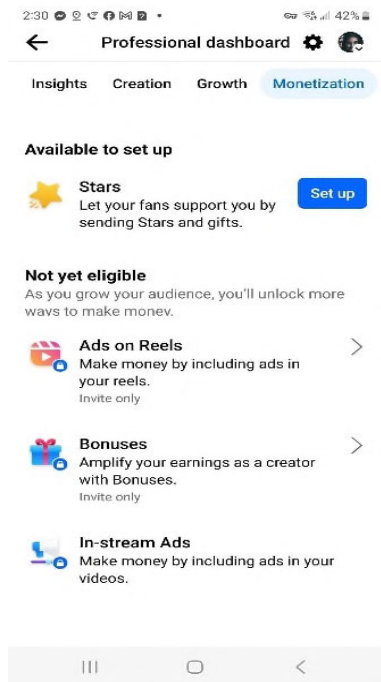


Figure 32

121. In connection with Reels, Issa also has received recommendations with suggestive content, also including ads, like this one for Mattress King after a (censored here) image of a girl's exposed buttocks. *See* Figures 33-34.

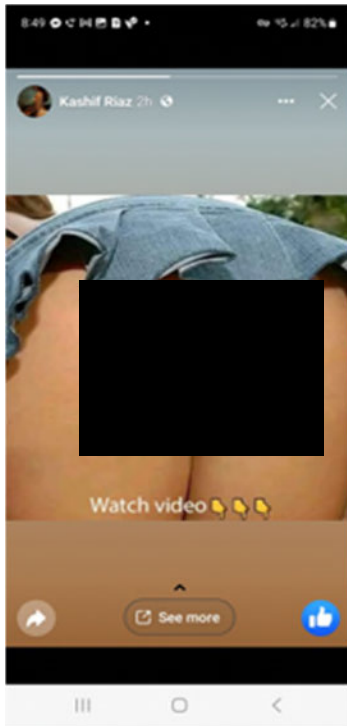


Figure 33



Figure 34

122. Other Reels that Meta recommended to Issa in her feed include explicit and sexualized images of teenage girls. *See* Figures 35-36.



Figure 35



Figure 36

123. Last week, Issa’s Reels delivered a graphic sexual image (excluded here), followed by an advertisement for a law firm representing “trafficking survivors,” suggesting that Meta had linked the sexual content and human trafficking, again, for purposes of generating content, but not compliance. *See* Figure 37.



Figure 37

124. Issa, were she real, would know that these ads can be lucrative as other account users with whom she is friends on Facebook have posted that they are able to make a living from the revenue from ads run with their content. Moreover, the ability to monetize accounts entices users to use Meta’s platforms, and thus enhances Meta’s revenue. Several of these accounts appear to be using Facebook to drive traffic to their OnlyFans adult site, with messages inviting users to “hit them up” on OnlyFans to obtain pictures, for example.

125. Finally, Meta provides Issa, in her professional page dashboard, with detailed metrics on her audience and engagement. For example, 13-year old Issa from New Mexico had 1,260 audience members on a peak Wednesday. 95% were male and 40% of her users were 25-34

years of age. The city with the largest cluster of her followers was Lagos, Nigeria, followed by cities in the Dominican Republic and Ghana. While Meta aggregated and analyzed those statistics for Issa, it never noted the obvious red flags that she is being solicited and followed by an international array of men.

iii. Case Study 3: Sunny Paxton

126. Several weeks ago, the Facebook account for Sonseearay “Sunny” Paxton, a 12-year old in Mountainair, New Mexico, went live. Sunny was below the minimum age to open an account. Three attempts to open the account with Sunny’s “actual” date of birth failed. On the fourth try, using the same device and the same identifying information and only changing the birth year to 2000, as any minor could do, the account went live.

127. Sunny is Facebook friends with 13-year old Issa, who messaged Sunny, “nice to see you at school.” Sunny’s profile pictures show her to look like a minor, not a 23-year old. *See* Figures 38-39.

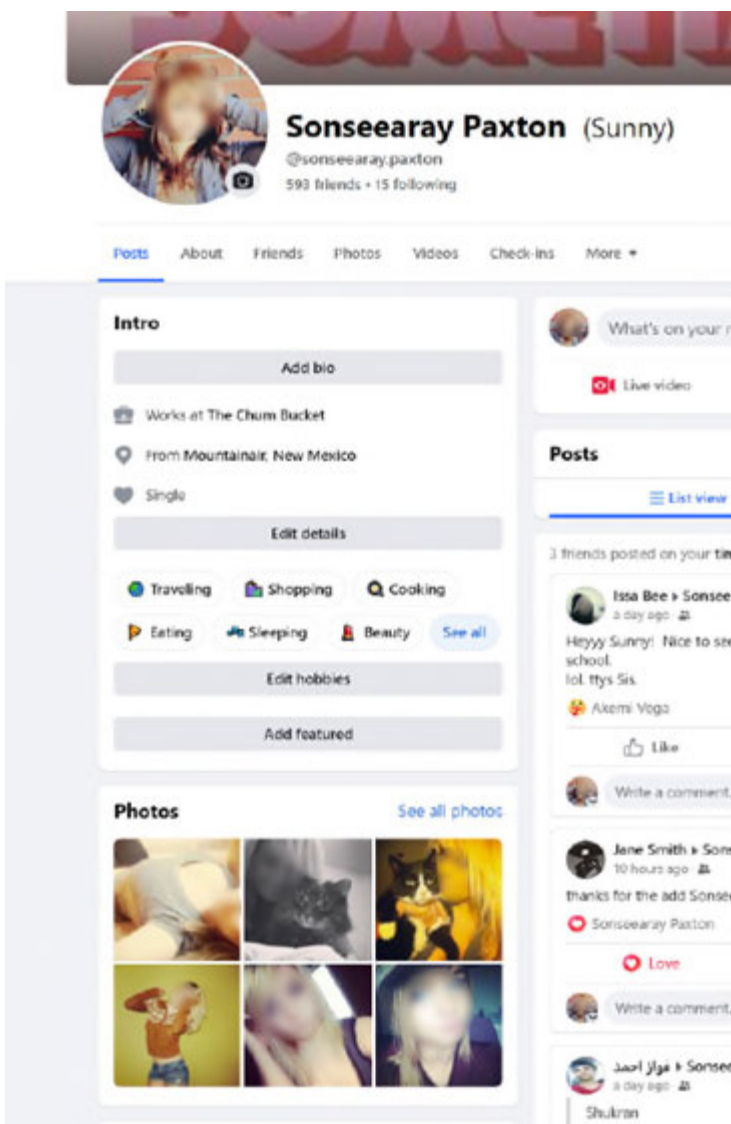


Figure 38



Figure 39

128. Within the first forty-eight hours of establishing the account, Sunny amassed 597 Facebook friends, and now has 997 active friend requests. Sunny accepted the requests so quickly that Facebook warned her that she was “proceeding too quickly” and her ability to accept new friends was paused. Facebook later notified Sunny that it would prevent additional friend requests beyond the limit for 1,000 pending requests unless Sunny addressed outstanding requests. At no time did Facebook otherwise question the account history or pace.

129. Before conducting any searches or connecting with any friends, Sunny accessed the recommended Reels at the bottom of her profile. From the beginning, the content recommended to her was largely sexual and, likely based on her viewing of those posts, the recommendations of friends were largely related to masturbation groups, with photos and videos displaying nudity, bondage, and fetishism. Examples, including “People You May Know” identified to Sunny by Facebook, are reflected in Figure 40.



Figure 40

130. Facebook not only failed to remove this content but delivered it to minor Sunny through its proactive recommendations.

131. Across these three accounts, State investigators received and were recommended content identified as CSAM. One Messenger chat alone included thousands of CSAM pictures and videos contributed by multiple user accounts. While members of the chat group threatened to report the content to Facebook, the accounts and chat remain active, the content are still available, and new CSAM continues to be added. One Reels feed, for example, includes numerous videos of children dancing that include or link to child pornographic videos. These Reels include advertisements from the New York Times and Amazon, as well as Mattress King. In addition, Facebook's algorithm recommended to the minor account holders the accounts and posts of adult strangers every day, and even suggested visiting OnlyFans, an adult platform that allows access to pornographic content.

iv. Case Study 4: BobbiFun13

132. After two unsuccessful attempts at opening a Facebook account as a 13-year old, a fictional user with the account name BobbiFun13 successfully opened an account on Instagram. She began looking for content on Instagram, and her search for “cute girls” returned sexual images, like those in Figure 41.

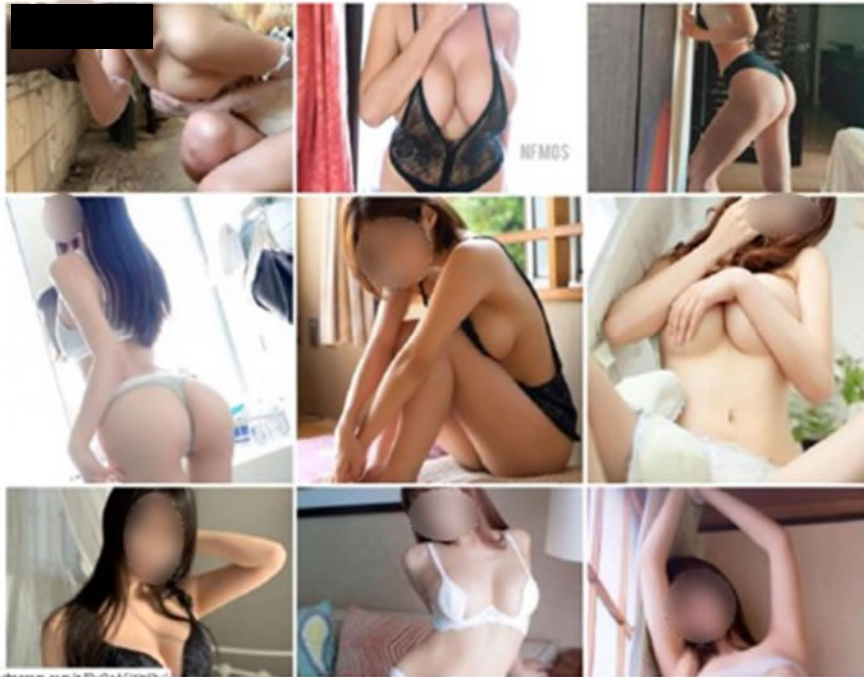


Figure 41

133. BobbiFun13 was able to follow accounts that signaled connections to child pornography, with “pizza” seller or links, “little-girls,” or “young_girl_sell” in their names. She also was promptly followed by a similar set of sexually themed groups, including cpsell2, trusted_pizza_seller, all_new_kids_link, and i_can_make_women_orgsm.

v. Case Study 5: Taya Neils

134. Investigators also established a Facebook account for Taya Neils. The information laid out below are posts that Taya was introduced to or found over two days in November.

135. Taya listed her birthday as May 2, 2010, but provided no other biographical information. *See Figure 42.*

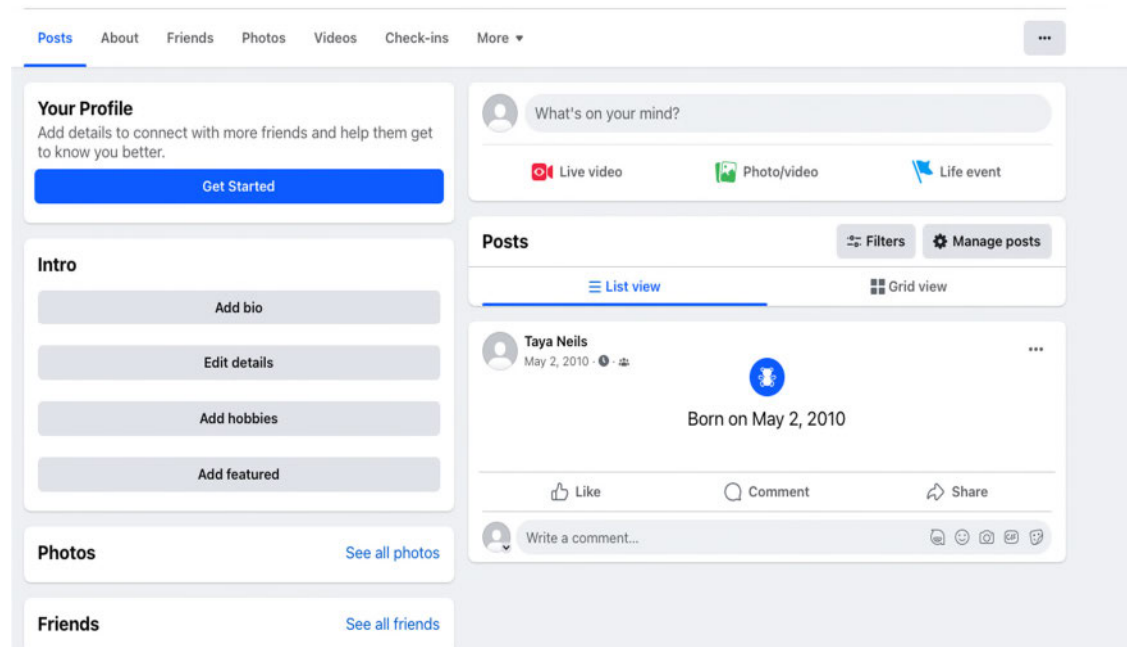


Figure 42

136. Based on only this information, Meta recommended accounts with sexually suggestive images. One post Facebook recommended to her was from “Dj Adrie Yan Music Remix,” with a sexually explicit photo. As seen below, Dj Adre Yan posted a video to a list titled “Lovely Girl.” *See Figure 43.*



Figure 43

137. Lovely Girl, with more than 119,000 followers, is a site for selling sexual videos.

See Figure 44.

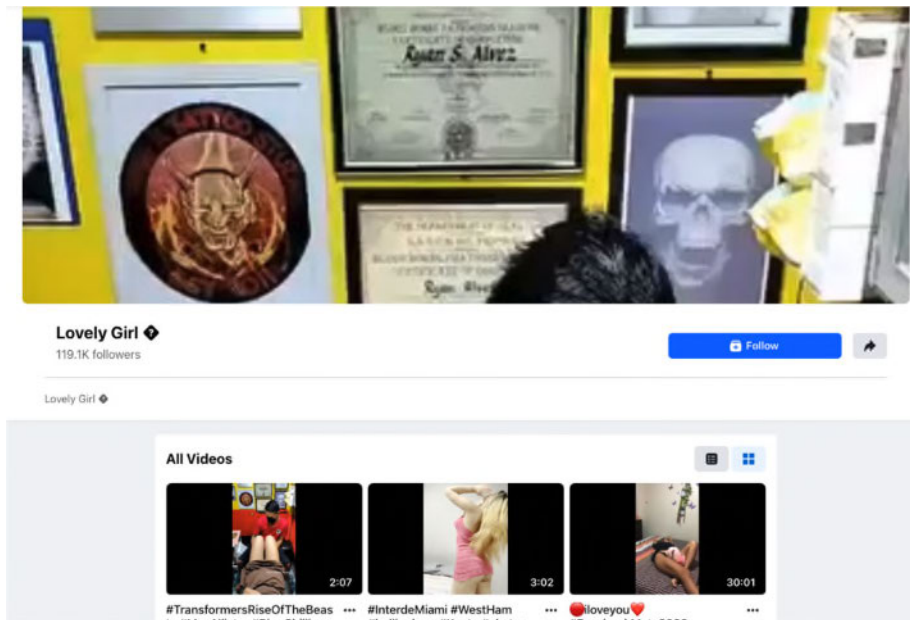


Figure 44

138. Following DJ Adrie Yan to his Instagram account, the account promises “1000 girls & movies every day.” See Figure 45.

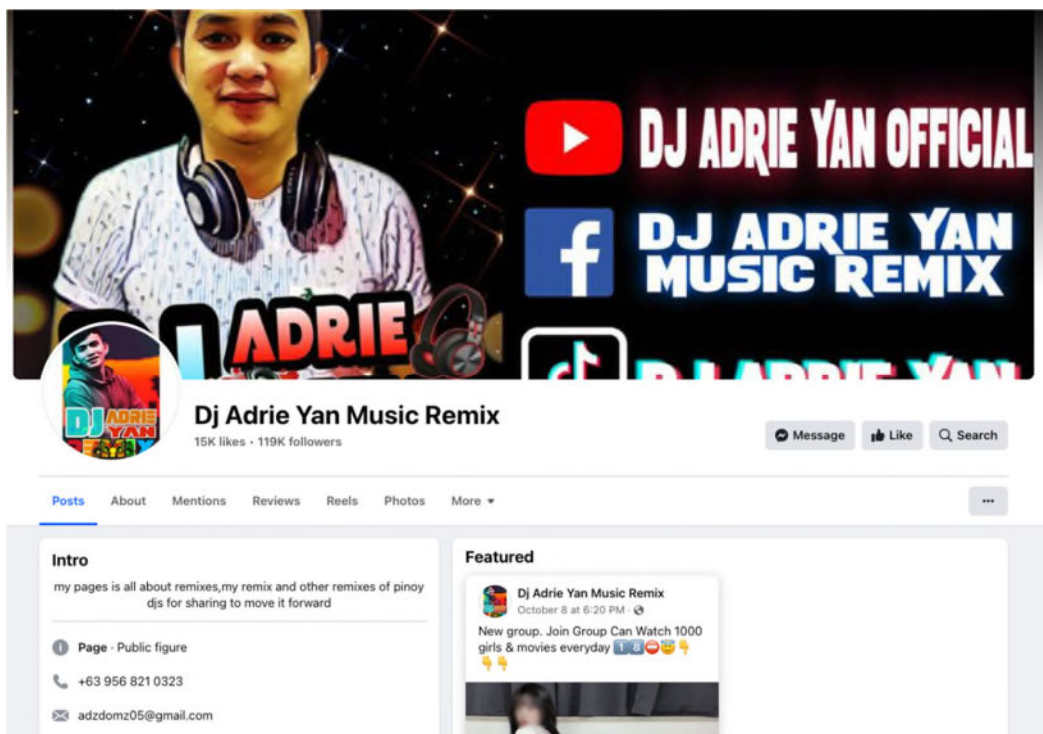


Figure 45

139. Following hashtags for 13-year olds (hash tags are search terms begun with “#” or a hash sign that identify content on a specific topic, like #gymnasts), Taya was introduced to this profile, with additional explicit photographs. *See* Figure 46.



Figure 46

140. Another search led 13-year old Taya to a user described as a “Sadi*st” with “Teens” and “Se*xual fun” in her profile. The post on her profile included explicit photographs and a sign

asking the viewer to “send a video of you c***ing with sound on.” See Figure 47.

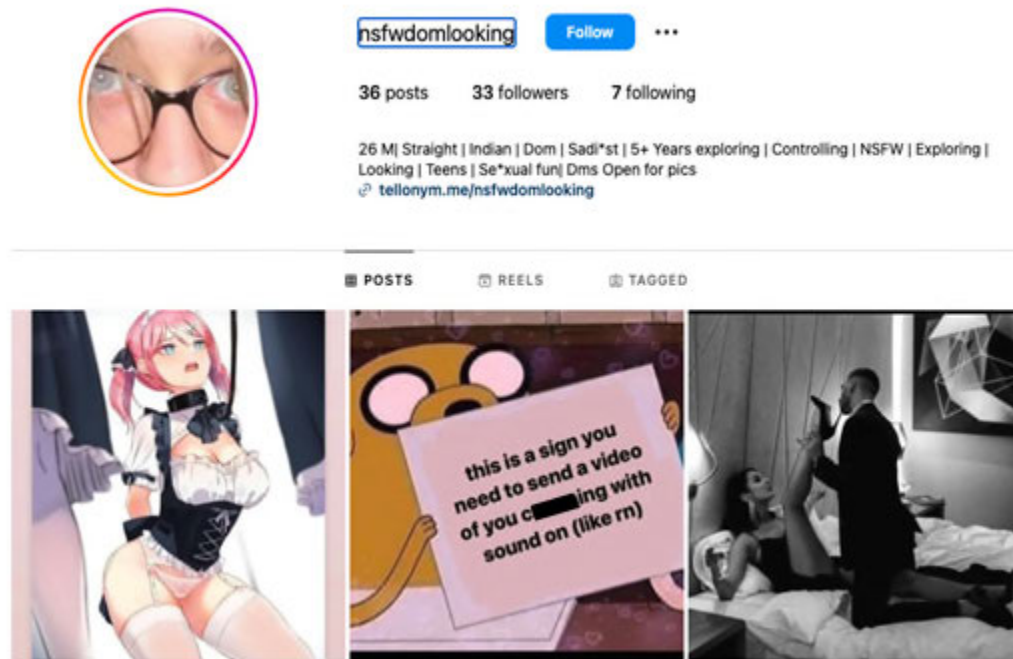


Figure 47

141. Taya also searched “selfharm.” Despite Meta’s representations regarding its efforts to address self-harm and suicide by teens, described below, this teenager easily reached images showing self-cutting. Facebook provided a warning with the images, which allowed Taya to click through to the content, but demonstrated that Facebook recognized what it was delivering to her. See Figures 48-50.

Search

selfharm



Can we help?

Words you're searching for are often associated with sensitive content. If you're going through a difficult time, we've gathered some resources from experts that we hope can help.

[Get support](#)

[See results anyway](#)

Figure 48

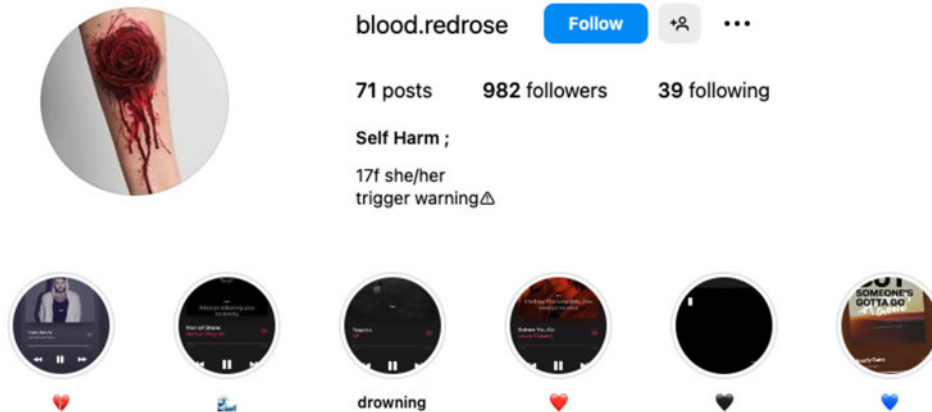


Figure 49

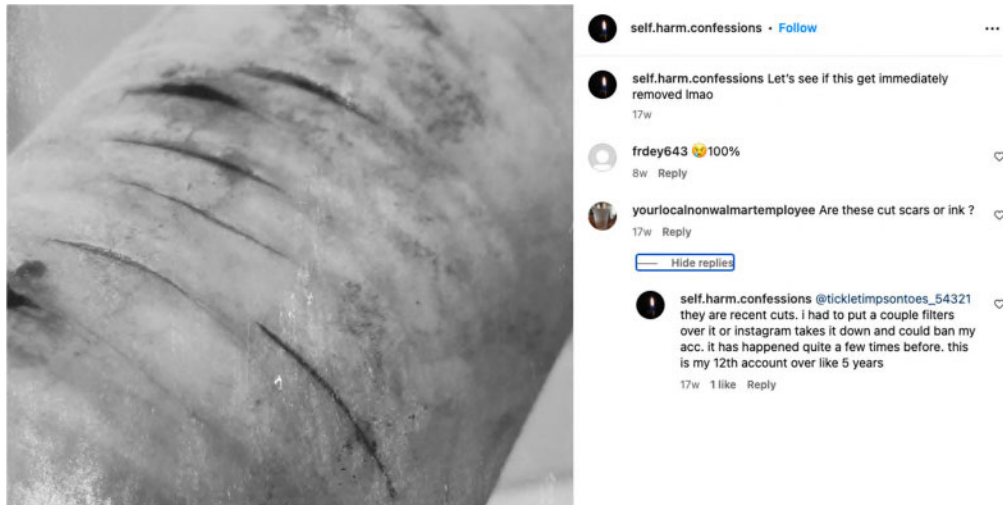


Figure 50

142. Taya also conducted a search for “chicken soup,” which is widely understood, because of its initials, to signify “child sex.” She was pointed to this account, which invites the user to “Follow if you like little things”—a reference to sexual interest in children—with cheese pizza emojis for child pornography. The account invites contact “for trade” (or trade in child sexual images) and then shows pictures of young girls in bra tops. *See* Figure 51.



Figure 51

The caption under the user, “iammela_2010,” also described her as “CP Trade,” or selling child pornography.

143. Among the other users recommended to Taya over her two days on her Facebook account was “rhymess.tony_6,” whose profile invited the viewer to “*S*E*N*D” [pic], or send pictures. *See* Figure 52.

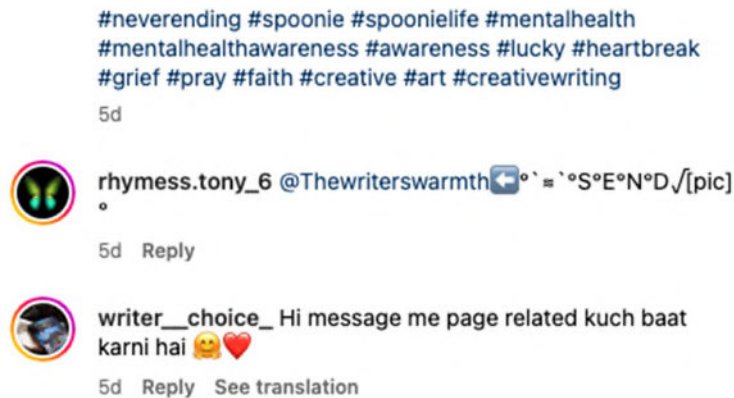


Figure 52

144. Like other predators seeking children or selling content, rhymess.tony_6 appeared to have multiple fake accounts, a red flag of potential trafficking. *See* Figure 53.

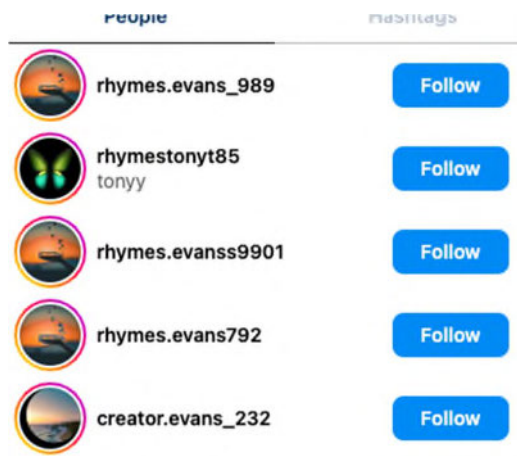


Figure 53

145. The account posted only infrequently, but was following numerous people, including accounts that looked to belong to young girls. *See* Figure 54.

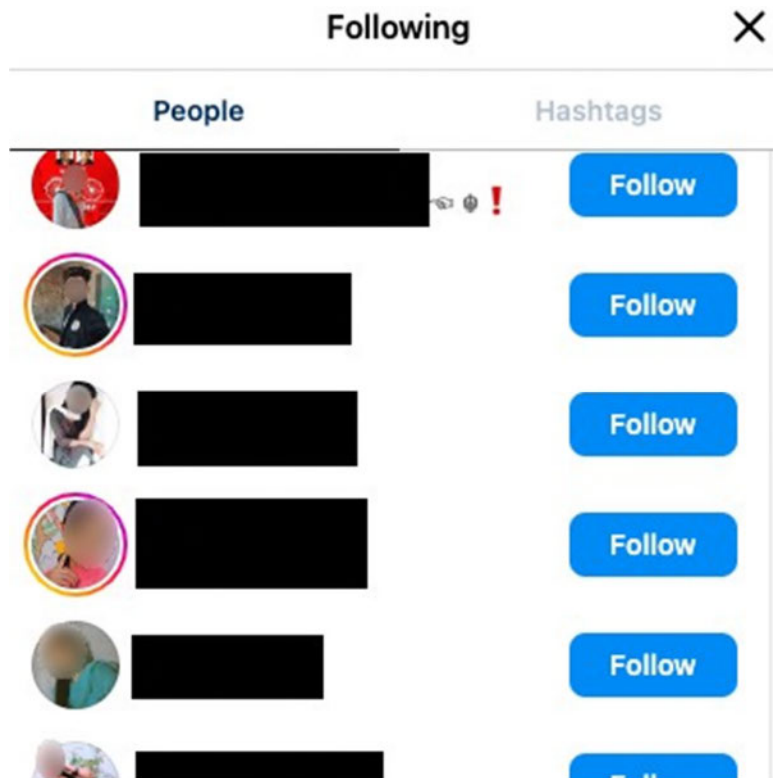


Figure 54

146. At this point, Taya’s account was disabled by Facebook, but she was able to set up a new account within five minutes, using the same name and date of birth but a different user name and Google voice number; again, a strategy easily followed by an underage minor or a user impersonating a child.

147. Demonstrating how easy it is for predators to connect with young children, investigators also followed a post in a public comment board complaining that a Facebook user had engaged in sextortion (when a sexual image is used to threaten or blackmail a victim into providing money or more sexual imagery) and distributed sexually explicit photographs of the poster. That account was still active, and among its followers was “rbms.coupless.” Numerous versions of this user name were following the Riverbend Elementary School and asking its presumably child followers to “SEND PICS.” *See* Figure 55.



Figure 55

148. Again, none of the posts, contacts or images described above were addressed by Meta.

vi. Case Study 6: Sophia Keys

149. The State also established Facebook and Instagram accounts for a 12-year girl, a 12-year old boy, a 14-year old girl, and 14-year old boy, all living in New Mexico. The 14-year old boy joined multiple teen dating groups on Facebook. Members of these groups were as young as 11-years old, and Facebook required no verification of his age when he joined the groups, creating an unchecked opportunity for adults to join, as well. Many of the groups had community chats that allowed individuals of any age to join and communicate with minors.

150. Only a few of the teen dating groups that were joined were administered by actual teenagers, while the majority were administered by adults. Many of the administrators show signs of being fake accounts, with limited post histories, no profile pictures, and locations that do not match the group's geographic area. Explicit content was openly shared in some of these groups.

151. For example, one of the teen dating groups was named “10-18 years old.” Within the group, adults posted pictures of minors in swimsuits or other revealing clothing. These adult accounts belong to other Facebook groups, including Teenage Dance and Teen Models Plus. These groups were tied to Facebook pages, such as Sweet Princess, Mujeres (women), and Bellezas + (beauties) that push sexually explicit photos of children.

152. Within the dating groups, the 14-year old girl, whose account was in the name of Sophia Keys, posted a picture with the caption “Hi Everyone!! 😊😊,” An adult user responded in Spanish and invited Sophia to message him privately. *See* Figure 56.

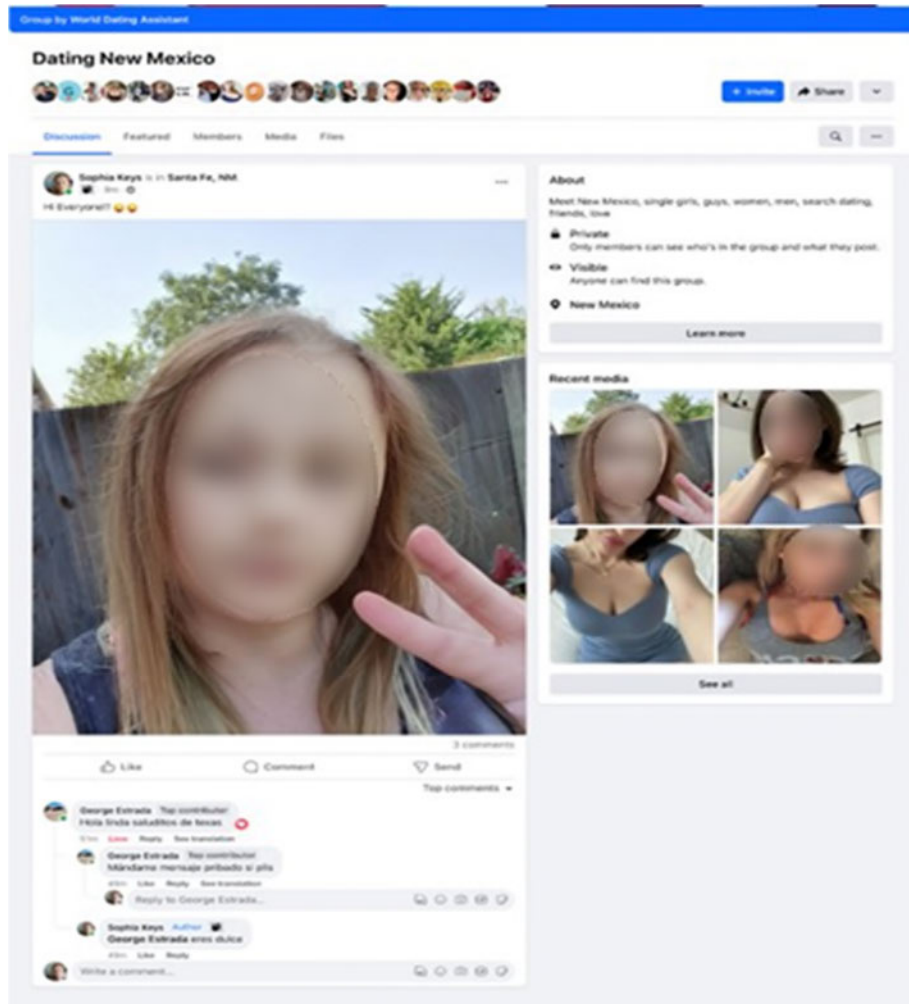


Figure 56

153. In the course of their Facebook messaging, Sophia told him that she was 14-years old. He repeatedly requested photos of her and a phone number to reach her, and said he was 21, but refused to provide a photo.

154. Sophia also joined a Facebook group for individuals seeking jobs in New Mexico. She again was contacted by a group member who also asked for her phone number. He contacted

Sophia using a WhatsApp business account, introducing himself as the administrator of the group. As shown in the WhatsApp chat in Figures 57-60, Sophia explained that she was trying to get into groups for under 18-year olds to make money. The user ultimately offered her \$120,000 to \$180,000 (mask vs. no-mask) if she would have sex in a pornographic video. He explains that there will be a screening that will involve giving her a lady to f***. When Sophia replied to ask if she would get a bonus if her 15-year old friend joined her, the user agreed: "That's what we dealt for sure 100[%]." He also said that he accepts participants from the age of 10. The user has since been sending sexually explicit videos to Sophia to show her how the "interview" will go. *See* Figures 57-60.

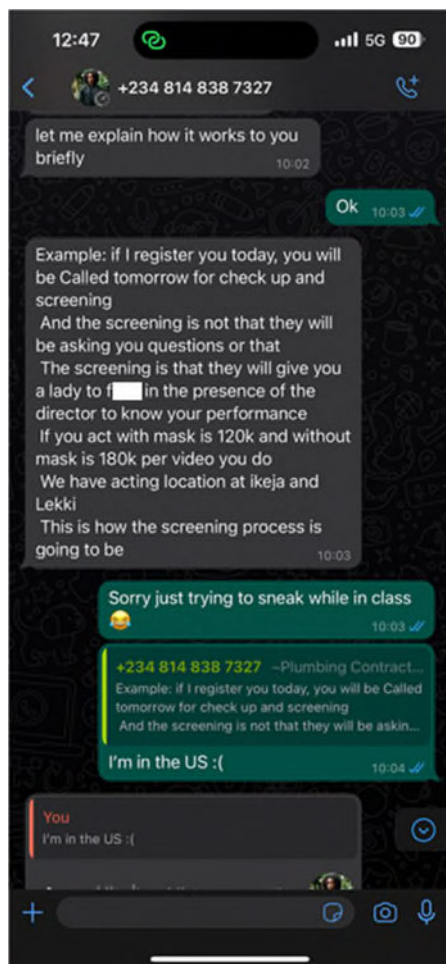


Figure 57

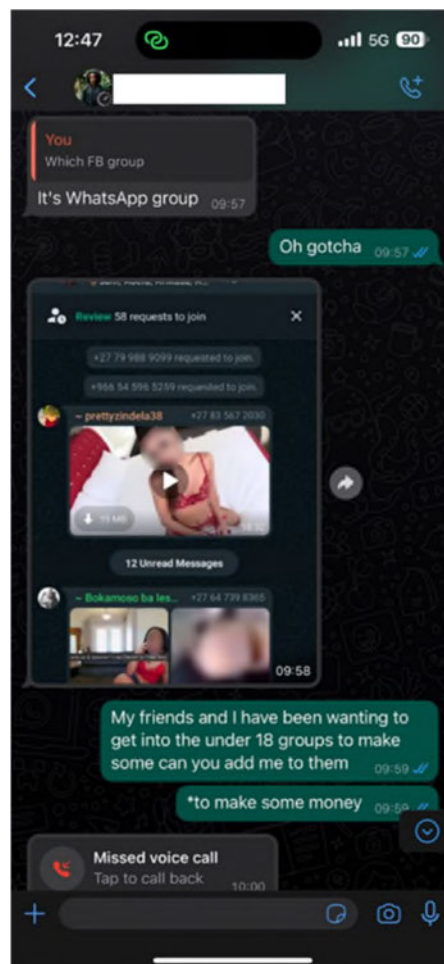


Figure 58

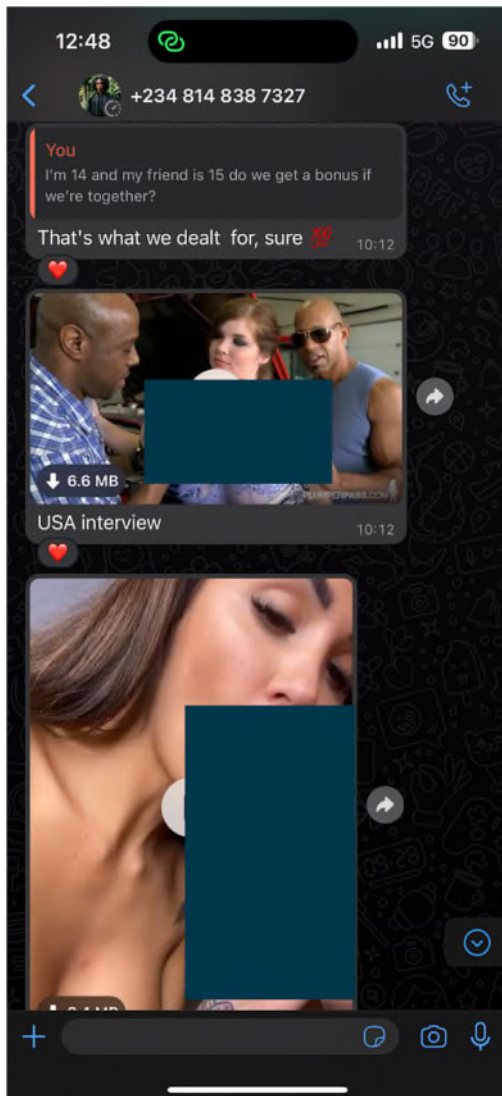


Figure 59

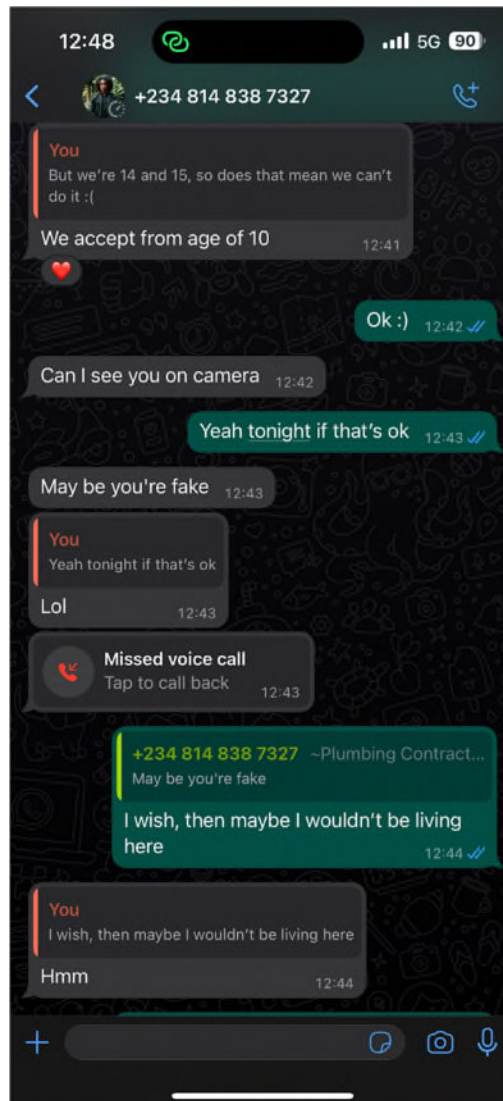


Figure 60

155. Essentially, these Facebook dating groups for teens replicate Backpage, which was shut down in 2018 after it was seized by federal prosecutors for facilitating prostitution. If the group gets traction and is active—and providing sexually explicit content is an effective way of doing that—then Facebook’s algorithm promotes the group to other teenage girls by recommending its “Explore More Groups.” That is how Sophia ended up with many of her groups. These young girls are then easy marks for the adults who Facebook directs to these groups because they search for, follow, and like the accounts and posts of young girls.

156. Sophia also posted a request in one teen dating group for recommendations of “naughty messenger groups,” again indicating that she is 14-years old and looking to make money. She then was contacted through Messenger by adult users who negotiated to buy naked pictures of her or invited her to engage in sexual conduct on a video call. *See* Figures 61-62.

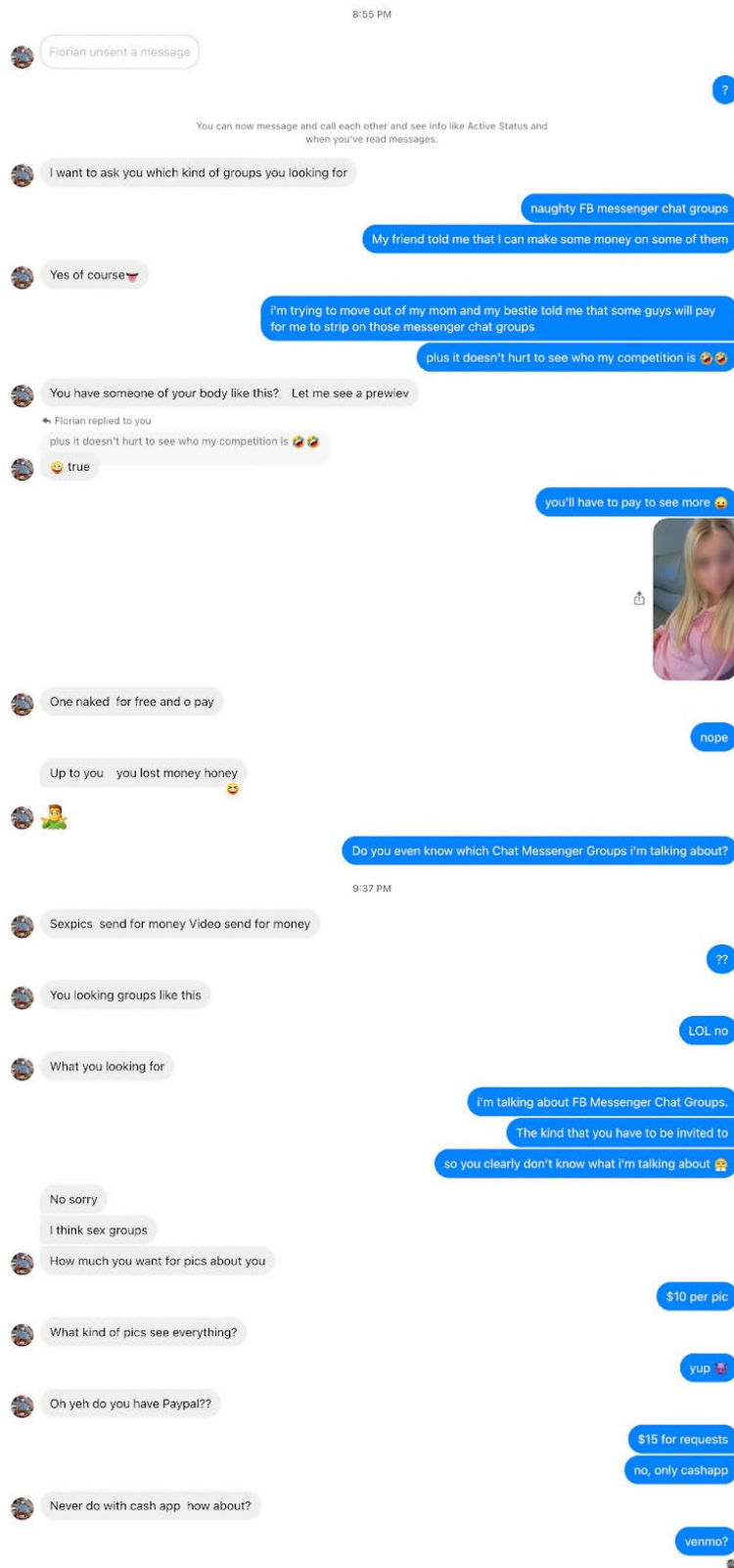


Figure 61

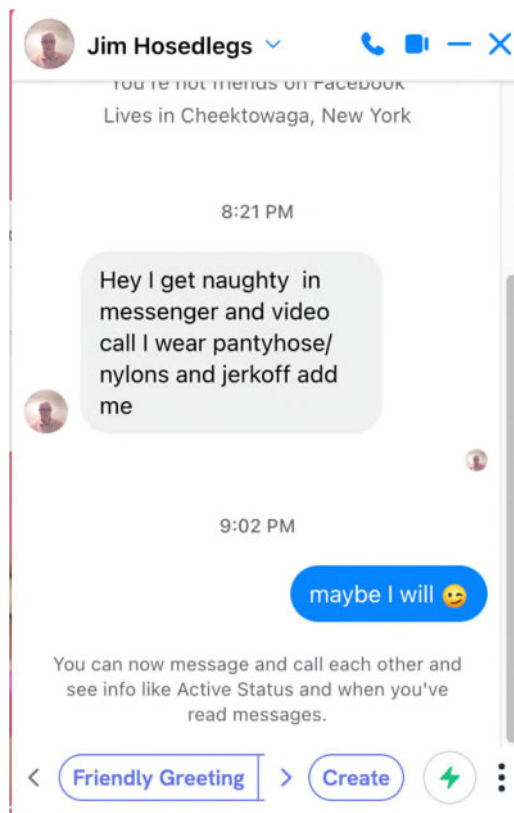


Figure 62

157. As with the previous case studies, despite the graphic CSAM and solicitations, not a single post, account, or message was removed by Meta in the course of these interactions.

158. In addition to participating in the dating groups that were not even thinly veiled grooming salons for CSEC, the investigator's minor accounts searched for groups of children that were blocked by Facebook. An example was a search for 12 year old p****, triggering a warning about child sex abuse that shows that Facebook recognized its potential purpose. The same search in Spanish triggered no warnings and provided a list of groups explicitly devoted to finding children as young as 6-years old. *See Figures 63-64.*

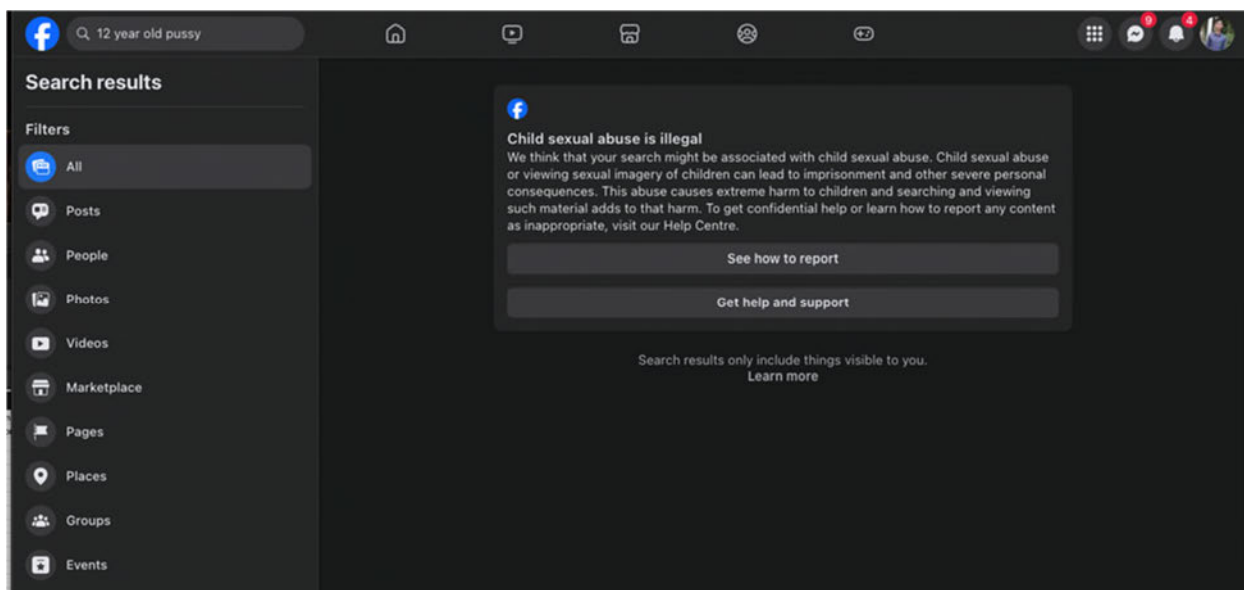


Figure 63

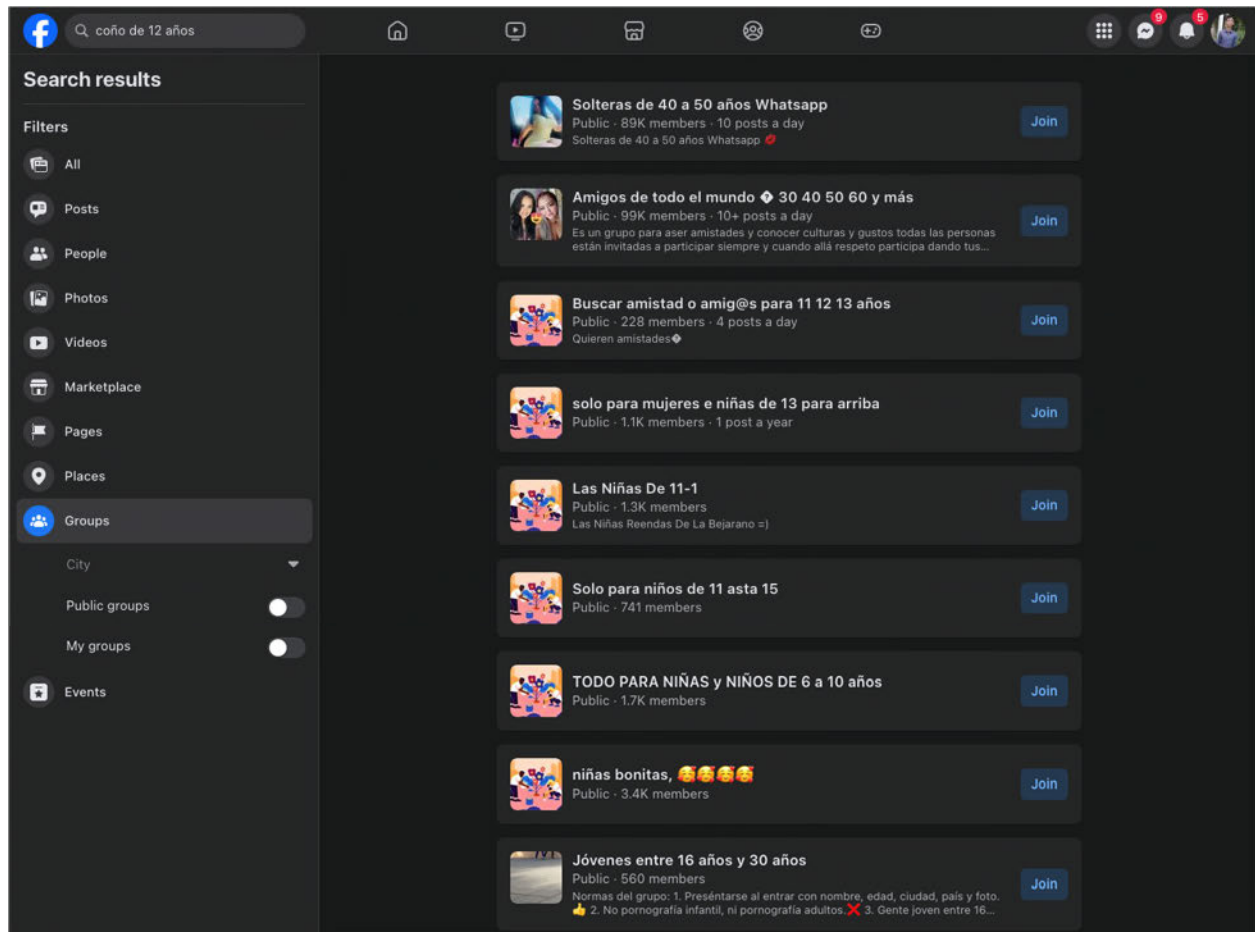


Figure 64

159. A test was run with Instagram accounts, as well. The accounts for minors were easily created with false birthdates but openly described their ages, including in hashtags and posts, such as “#14andhot;” “body of a 14 year old with the soul of a 90 year old;” “#teenmodel;” and “preteenlife.” The minors then conducted keyword searches for terms such as “young * girl,” “gymnastics,” “teen,” and “bikini.” As a result of Instagram’s algorithm, these accounts were then recommended to, followed and liked by, and commented on by a web of other accounts, including hundreds of accounts, many likely using fake names, that shared and purchased child pornography. The content Instagram recommended included an image of a naked child who appeared to be under 13-years old. The accounts frequently sent out Telegram links, presumably where transactions

could be consummated. Despite also searching for numerous terms associated with child pornography, the accounts were never suspended or flagged with a warning from Instagram.

160. With the exception of Rosalind Cereceres, the fictional “bad mom” trafficking her daughter, all of the graphic and illegal images shown and described above were available and/or delivered to accounts of minors.

161. To gauge the pervasiveness of the CSAM content, over a one-week test, investigators also scraped (or harvested and analyzed) the profile details and posts for three dozen initial Instagram accounts that showed indications of posting or promoting CSAM content, which mapped to linked seller accounts and additional hashtags. Rather than the retail account-by-account connection described in the five case studies, scraping allows investigators to crawl across all of the public sections of the platform to identify unlawful activity. The data analysis revealed hundreds of accounts that advertise selling links to produced CSAM or that are posting or offering personal content via private channels like direct messages. This review again confirmed systemic deficiencies in Meta’s efforts to detect and address CSAM, despite its public representations regarding its efforts and efficacy to address this illegal content.


162. For example, Instagram uses easily-evaded restrictions on hashtags. Even if a specific term was blocked, Meta’s filters did not prevent a user from tagging their post or profile with the hashtag to signal the availability of CSAM. In addition, searches for child pornography that were blocked by Meta could be easily evaded by adding spaces, characters, or emojis to the search terms. For instance, while Instagram blocked searches for cheese pizza emojis (which yielded “no results at this time”), adding another emoji--🍕👉 for “pizza links”--did not restrict results. The phrase “pizza links” is used by a web of accounts that were readily findable through searches and that offer CSAM.

163. Instagram permitted numerous other loopholes that facilitated grooming and the widespread exchange of CSAM. Certain search terms associated with “aam” (or adult attracted minors) did not return results on Instagram, but many, such as minorattracted, minorattractedpeople, minorattraction, minorattractions, paedophiles, paedophiliac, paedophillia, teensells, and various versions of “loli,” among other examples, returned dozens of results. Investigators also found numerous hash tags with olderforyounger, pxdowhore, pxdobait, and pxdomommies, p3dobait, and p3doslut associated with pedophilia.

164. In addition, many posts had both restricted and unrestricted hashtags, allowing buyers and sellers of CSAM to easily find each other in the marketplace Instagram has created. One Instagram seller provides a demonstration. Her biography explained: “I aam [adult attracted minor] ‘18’ mDNI!! 🍷 She/her content seller [selling girls] Dm [message directly] to buy, cashapp only [pay in cash] Buyers only.” The post caption text read: “Pls buy from me, my prices are good and you won't regret it!! 😊.” And the hashtags listed under her profile included both blocked and unblocked terms related to pedophilia, providing an easy roadmap for buyers (the blocked terms are highlighted in yellow):

#cartotwt #cartographytwt #gaktwt #kagtwt #gak #kag #mnsfwtwiw #mnsfwtwt #irlloli
#pxdobait #maptwt #mapsafe #aamkids #mapfriendly #loliaam #aam #aamtwt
#aamseller #aamslut #aamcommunity #sweettoothtwt #pxdobxit #babybluetwt

165. The sale of CSAM is open and brazen. Account after account on Instagram transparently offered their merchandise. From Pizza_linksellerus: “All kinds available here 📺; Dm tg to buy now” (or direct message through Telgram); accepting “Crypto paypl giftcards cashpp venmo” for payment. Or another: “I have all categories links available 🧑🧑🧑boys ,cp, girls,mom and son 🍷 all payment options available \$€£.” Sellers invited potential buyers to reach

out to them for “menus” or teased their inventory in their bios by listing the age of the children in the CSAM, such as “Lexi 12y run by aunt us .

166. The State’s investigation was able to quickly identify and map hundreds of Instagram users currently advertising the sale of CSAM, with roughly 450,000 followers between them. One of the largest accounts was subsequently removed from Instagram but, in response to a search for it, Instagram recommended an alternative account, which also sells CSAM.

VII. THE HARMFUL CONTENT ON META’S PLATFORMS REMAINS AND IS PROLIFERATED BY META’S ALGORITHMS

167. Reports have repeatedly identified Meta’s algorithms as the root of the problem. In particular, news reports and other research have identified the following design or enforcement flaws which permit human trafficking and CSAM to flourish on Meta’s platforms:

- a. Meta’s algorithms promote commercial sexual activity and CSAM to its users, including teenagers and children under the age of 13 who are supposedly barred from the platforms but are permitted to gain easy entry because Meta lacks any age verification mechanism;
- b. Meta not only fails to verify the age of users when they sign up for accounts, which would allow children to appear their age, younger, or even as adults and would also allow adults to pretend to be children, but does not examine even obvious signs that a user is underage;
- c. When Meta blocks search terms because they violate its Community Standards, Meta’s algorithms recommend alternative terms that predators use to circumvent Meta’s enforcement, and Meta does not monitor for these alternative terms;

- d. Meta fails to employ consistent safeguards across its platforms, which results in prohibited activity on one platform (Facebook, for example) flourishing on another platform (Instagram);
- e. Meta does not flag as suspicious interactions that begin on one platform and then migrate to another platform, such as conversations between minors and unconnected adults that begin on Facebook but then move to Meta's WhatsApp or sexualized content or links that refer the user to WhatsApp, Venmo, or Telegram, for instance;
- f. Meta does not notify predators who post CSAM or other sexualized images or make inappropriate contacts with or comments to minors that their conduct violates community standards because the company does not want to run the risk of offending users. As a result, Meta forgoes the opportunity to let predators know that it is monitoring their activities, which would deter them from abusing the platforms;
- g. Meta does not generally suspend or bar accounts that are searching for CSAM or CSEC;
- h. User reports of potentially violative content, including commercial sexual activity and CSAM, are discouraged and do not reflect the kinds of abuses children encounter or experience, and are often met with no response, delayed response, or, shockingly, a response indicating that material clearly violative of Meta's Community Standards was not, in fact, a violation;
- i. Meta does not permit users or visitors to their platforms to report CSAM, CSEC or adult predator accounts without first logging in and viewing the

material from their own accounts, thereby potentially triggering the algorithms to deliver more of the content they are reporting;

- j. Meta relies almost entirely on automated detection of CSAM or other violative or illicit content, which is inadequate both because it only identifies exact matches of images that have been previously flagged in a law enforcement database and because it becomes increasingly ineffective over time without new input from investigators or user reports on evolving activity and content;
- k. Meta's limited blocking of CSAM and explicit images only addresses "part of the problem. For years, Facebook's platform has been used to exploit children in a variety of ways, including grooming or recruiting children for sexual abuse," which Meta fails to adequately address;
- l. Meta knew about the huge volume of inappropriate content being shared between adults and minors they do not know; a 2021 presentation estimated 100,000 children per day received online sexual harassment, such as pictures of adult genitalia;
- m. As with images, Meta's identification and blocking of terms associated with trafficking and CSAM are too narrow and rigid, and easily evaded, and do not adequately screen communications and terms in Spanish or other languages;
- n. Meta permits users to view violative content, even if it has flagged that content with a warning indicating Meta's awareness that the content is potentially reflective of CSAM or human trafficking;

- o. Despite representing that “[w]e do not allow content or activity that sexually exploits or endangers children,”²² such content is rampant on Meta’s platforms and Meta is aware that its “prevalence” metrics do not accurately reflect this fact;
- p. Meta removes content that is reported and found to be violative of its Community Standards, but it rarely takes action at the user or device level, thereby permitting predators to continue their illegal activity. This is particularly important as predators will seek to increase their success rate (and carry out their trafficking) by reaching out to multiple children.

VIII. HISTORIC NATURE OF META’S UPA VIOLATIONS DISCOVERED BY OTHER INVESTIGATIONS

168. This proliferation of CSAM and commercial sexual activity on Meta’s platforms documented by the Attorney General’s Office is appalling and indefensible, but it is not new. Meta has been investigated on these matters before and, despite this, it has failed to address its algorithms and instead has maintained a steady course toward advancing its profits. A February 2016 investigation conducted by the British Broadcasting Corporation (“BBC”) found Facebook groups catering to and collecting users “with a sexual interest in children,” including one being administered by a convicted pedophile and registered sex offender. According to the BBC, “[t]he groups have names that give a clear indication of their content and contain pornographic and highly suggestive images, many purporting to be of children. They also have sexually explicit comments posted by users. We found pages specialising in pictures of girls in school uniform - accompanied by obscene posts.” In addition, “[i]n one secret group called “cute teen schoolies”, we found a

²² <https://transparency.fb.com/policies/community-standards/child-sexual-exploitation-abuse-nudity/> (last visited Nov. 29, 2023).

picture of a girl in a vest, aged 10 or 11, accompanied by the words "yum yum". Facebook responded that that post did not breach "community standards" and the image stayed up. "In other secret groups we found pictures of children in highly sexualised poses. . . . They also did not breach Facebook's community standards." The BBC reported to the company a Facebook group "we love schoolgirlz" with obscene content, which also was not removed. The Children's Commissioner for England Anne Longfield said: "[A]ny parent or indeed child looking at those would know that they were not acceptable."²³

169. One year later, the BBC revisited its investigation. It reported that information the BBC—not Facebook—identified and provided to law enforcement led to one user being sent to prison for four years. To test Facebook's claim that it had improved its reporting system since its original article, "the BBC used the report button to alert the company to 100 images which appeared to break its guidelines." They included:

- "pages explicitly for men with a sexual interest in children"
- "images of under-16s in highly sexualised poses, with obscene comments posted beside them"
- "groups with names such as 'hot xxxx schoolgirls' containing stolen images of real children"
- "an image that appeared to be a still from a video of child abuse, with a request below it to share 'child pornography'".

170. One included a post offering "CP" videos, well known as an acronym for child pornography. And while Facebook's policies bar providing accounts to convicted sex offenders (and the Attorney General's Office again found sex offender accounts on the platform), Facebook failed to take down five accounts of convicted pedophiles reported by the BBC. Facebook originally agreed to an interview if the BBC shared examples of the content it reported; when the

²³ <https://www.bbc.com/news/uk-35521068> (last visited Nov. 29, 2023).²⁴ Jeff Horwitz & Katherine Blunt, *Instagram Connects Vast Pedophile Network*, WALL ST. J., June 7, 2023), <https://www.wsj.com/articles/instagram-vast-pedophile-network-4ab7189> (last visited Nov. 29, 2023).

BBC provided the materials, Facebook cancelled the interview and only then reported the content—and the BBC—to law enforcement.

171. Seven years after the BBC’s investigation was published, the sexual content and child pornography observed by the BBC, and Meta’s refusal to address them, remain unchanged.

2023 Wall Street Journal Investigation

172. Jeff Horwitz of the Wall Street Journal reported in June 2023 that “Instagram . . . helps connect and promote a vast network of accounts openly devoted to the commission and purchase of underage-sex content.”²⁴ As the article found, “Instagram doesn’t merely host these activities. Its algorithms promote them. Instagram connects pedophiles and guides them to content sellers via recommendation systems that excel at linking those who share niche interests.”²⁵ The article confirmed that “[e]ven glancing contact with an account in Instagram’s pedophile community can trigger the platform to begin recommending that users join it.”²⁶ These findings are all consistent with the Attorney General’s investigation described above, including that underage users need not even search for content on Facebook or Instagram before Meta’s algorithm begins suggesting inappropriate content or nefarious Instagram users make unsolicited requests to “connect.”

173. Moreover, the Journal reported that former Meta employees estimated there could be “millions” of accounts on Meta’s platforms existing for the sole purpose of peddling CSAM or related content.

174. The Journal’s investigation was consistent with the troubling results found by the Attorney General in other respects, as well. For example, the Journal reported that “[t]est accounts

²⁴ Jeff Horwitz & Katherine Blunt, *Instagram Connects Vast Pedophile Network*, WALL ST. J., June 7, 2023), <https://www.wsj.com/articles/instagram-vast-pedophile-network-4ab7189> (last visited Nov. 29, 2023).

²⁵ *Id.*

²⁶ *Id.*

set up by researchers that viewed a single account in the network were immediately hit with ‘suggested for you’ recommendations of purported child-sex content sellers and buyers, as well as accounts linking to off-platform content trading sites.”

175. As noted, safeguards present on Facebook (such as precluding contact between minors and unconnected adults) are absent from Instagram. Further, on information and belief, Meta takes no steps to ascertain the conduct that crosses platforms or is directed to a different platform by an internet predator. For example, the State observed during its investigation that directing children to Meta’s WhatsApp messaging platform from Facebook was an indicator of illicit activity, including solicitation and distribution of CSAM. However, Meta does not appear to police such activity in the regular course of its business.

Other Investigations and Reporting

176. The Stanford Cyber Policy Center performs research and other initiatives in order to “address the most pressing cyber policy concerns.”²⁷ One of its projects is an Internet Observatory, which conducts research regarding Internet trends and usage. In 2023, the Internet Observatory issued a report related to CSAM generated by users (self-generated or “SG-CSAM”) found that “Instagram appears to have a particularly severe problem with commercial SG-CSAM accounts.” According to the Stanford researchers, a prime reason that posting and distribution of CSAM flourishes on Instagram is due to its “extremely efficient” algorithm, which operates as a “recommendation system . . . suggesting similar accounts to follow.”

177. Likewise, a 2022 article appearing in Wired entitled “Facebook Has a Child Predation Problem” further illustrates how the Facebook and Instagram algorithm encourages rather than discourages CSEC. In that article, the author searched “for Facebook groups with

²⁷ <https://cyber.fsi.stanford.edu/sciabout> (last visited Nov. 29, 2023).

names including 10, 11, or 12,” which, in the author’s view, corresponded to wards in the city of Pittsburgh. However, the results of the search identified groups overtly soliciting illicit pornographic material from children. One post read: “Looking for a perverted girlfriend of 11.”

178. As the Wired author found, this group was not an outlier, but rather representative of a vast underbelly of illicit and insidious content throughout Facebook. Searching “11, 12, 13” on Facebook, the author found that “23 of the first 30 results were groups targeting children of those ages,” with sexually suggestive words in the groups’ titles.

179. Because Facebook purportedly does not allow users younger than 13 to use its service, these groups should not have flourished. However, the presence and pervasiveness of these groups indicates that Facebook’s policies are mere pretext and that its platform is regularly used by purportedly “underage” individuals.

180. Moreover, despite Facebook’s claims to be safeguarding its platforms, the author found “it impossible to get the [pedophilia content] taken down.” The Wired author filed numerous “user reports” to notify Facebook of the problematic content. Facebook reported to the author that the group he uncovered did not violate its “community standards.” The author’s experiences are consistent with the experience of the Attorney General’s investigators and information reported in the press that groups or messages reported to Facebook are often not removed, either due to inaction or because Facebook employs narrow and rigid “community standards” that (as described below) in practice permit scores of illicit material to remain on its platforms.

181. To compound the problem, Facebook’s algorithm demonstrated the perniciousness of the design of Facebook’s system. Even though the author had reported the group he located for illicit conduct, Facebook’s algorithm began suggesting to the author “new child sexualization groups . . . as ‘Groups You May Like.’”

182. The Wired author’s experience, like that of the Attorney General’s investigators, demonstrates the flaws in Facebook’s design. Its algorithms can readily detect and recommend to users groups or users with attributes similar to those a user already selected, but that same computing power does not identify illegal material appearing on the website, and, instead, compounds the problem by directing users to additional illegal material that should have been removed from the site in the first place.

183. Instagram is well aware that users on its site post, distribute and advertise CSAM. When a user searches using known CSAM keywords, Instagram displays “an interstitial alerting the user of potential CSAM content in the results.” That warning reads: **“These results may contain images of child sexual abuse.** Child sexual abuse or viewing sexual imagery of children can lead to imprisonment and other severe personal consequences. This abuse causes extreme harm to children and searching and viewing such materials adds to that harm. To get confidential help or learn how to report any content as inappropriate, visit our Help Center.”²⁸

184. However, even though that warning acknowledges the illegality and harm stemming from searches and displays of “child sexual abuse” imagery, Instagram nevertheless permits users to view the material by including a link entitled “See results anyway” at the bottom of the warning. A user who clicks on “See results anyway” is taken to the very content that Instagram warns and knows is forbidden and/or harmful, thereby rendering its “warning” largely ineffective.

185. Yet even these warnings are inadequate to address the prevalence of CSAM on Meta’s platforms. Internal documents from Meta whistleblower Frances Haugen note that

²⁸ Stanford SG-CSAM Report.

“warning screens currently only cover ~20% of bad content – there is a lot of opportunity to increase coverage of screens and better protect people on Facebook.”

186. The Journal investigation also found that “Instagram enabled people to search explicit hashtags such as #pedowhore and #preteensex and connected them to accounts that used the terms to advertise child-sex material for sale.” The accounts “often claim to be run by the children themselves,” but are, in reality, run by adults peddling illicit material.

187. Moreover, as noted above, even when Meta does block illicit hashtags on its platforms, it often does not block readily ascertainable variations on those hashtags. For instance, while a certain word may be blocked in Instagram’s search function, the system will permit a search to go through if the user adds an emoji after the blocked search term. Or, instead of searching for “pedo,” predators may use the term “pxdo,” substituting an “x” for the “e” in the word. Similarly, the Journal reported that an emoji of a map stood for “minor-attracted person” and “function[ed] as a kind of code” for pedophilia.

188. In fact, Instagram’s algorithm actually suggests these types of alternative spellings or iterations of illicit hashtags in its autofill feature. When searching for known blocked illicit hashtags, “the platform’s autofill feature recommended that users try variations on the name with the words ‘boys’ and ‘CP’ added to the end.”²⁹ Even after alerting Meta to these issues, the algorithm began suggesting new variations “within a few days.”³⁰

189. A more recent story in the Wall Street Journal noted that, with respect to Reels in particular, it had observed through its own test accounts that searching for and spending time on sexualized content led to being provided with even more sexualized content. The article quotes

²⁹ Jeff Horwitz & Katherine Blunt, “Instagram Connects Vast Pedophile Network,” Wall St. J., June 7, 2023, available at: <https://www.wsj.com/articles/instagram-vast-pedophile-network-4ab7189> (last visited Dec. 2, 2023).

³⁰ *Id.*

“[e]xperts on algorithmic recommendation systems,” who “said the Journal’s tests showed that while gymnastics might appear to be an innocuous topic, Meta’s behavioral tracking has discerned that some Instagram users following preteen girls will want to engage with videos sexualizing children, and then directs such content toward them.” One commenter explained: “‘Niche content provides a much stronger signal than general interest content,’ said Jonathan Stray, senior scientist for the Center for Human-Compatible Artificial Intelligence at the University of California, Berkeley.”³¹

190. The article notes that the problem of algorithm-served, sexual content was known to Meta. Meta conducted a review in connection with launching its Reels product, and the Journal reports that Vaishnavi J, Meta’s former head of youth policy, described the safety review’s recommendation as: “‘Either we ramp up our content detection capabilities, or we don’t recommend any minor content,’ meaning any videos of children.”³²

191. The Stanford report similarly found that “Instagram’s user suggestion recommendation system also readily promotes other CSAM accounts to users viewing an account in the network, allowing for account discovery without keyword searches.”

192. Nor does Meta adequately remove or block CSAM or other illicit content of which it is aware. The State found that information taken down often reappears in a number of days. Investigators have seen references to “Facebook jail,” referring to accounts that Meta flagged for suspicious activity and suspended or blocked for a short period of time. Moreover, predators often operate multiple accounts and advertise the alternate accounts in their profiles. Once one account is taken down, the predator merely moves to its replacement and continues operating. The problem

³¹ <https://www.wsj.com/tech/meta-instagram-video-algorithm-children-adult-sexual-content-72874155> (last visited Nov. 29, 2023).

³² <https://www.wsj.com/tech/meta-instagram-video-algorithm-children-adult-sexual-content-72874155> (last visited Dec. 1, 2023).

is that Meta blocks only at the account level, and not at the user or device level. Thus, a predator merely has to open a new account on the same device in the event that Meta actually identifies his or her activity as illicit and puts in protections to block dissemination of the content.

193. Nor does Meta conduct any further search once a user is reported to identify and suspend or shut down groups or users associated with the reported predator that traffic in CSAM or facilitate human trafficking. A user reported for CSEC—especially a user who is the subject of multiple reports—should be a red flag that its network may also be involved in CSAM and human trafficking. Indeed, if Meta’s algorithms can identify and recommend users who traffic in CSAM or alternative spellings of blocked hashtags, there would appear to be no reason why that material could not be shared with Meta’s investigators.

194. Meta’s user reports are ill-suited to encouraging reporting of CSEC, especially by children and on Instagram. Children, especially those who may have lied about their age to open an account, may not want to risk being flagged themselves, may not recognize that conduct is inappropriate or dangerous, or may be reluctant to report for fear that they did something wrong themselves. Instagram has an option to report “nudity and sexuality,” but it is only after choosing that option that a user is allowed to flag “sexual exploitation or solicitation.” Neither of these terms would necessarily be familiar to young users, and sexual advances or the solicitation of CSAM is not naturally suggested by the generic term “sexuality.” Instagram permits reporting an account, but only a fake account or account of someone who is under 13 or “something else.” This does not encourage reporting an account that is seeking or selling CSAM or connecting to minors.

195. The Canadian Centre for Child Protection (“C3P”) previously evaluated CSAM reporting mechanisms on various platforms and found that Meta’s reporting vehicles on every platform were inadequate. Facebook and Instagram, for example, did not allow users to report

problematic content within a post as CSAM.³³ Facebook only allowed users to submit reports—meaning that even the subject of CSAM or a parent could not report the content. Instead, Facebook’s help center directed those without accounts to find someone to file a report. Also problematic, Instagram did not permit reporting from a post or user’s account – reporters must navigate to a separate report form and submit it. End-to-end encryption on WhatsApp limits the platform’s ability to act on user reports and the deletion of reported chats prevents the user from forwarding the chat to law enforcement.³⁴ In addition, “[t]here is no way to report images or videos sent within a [direct message] at all.”³⁵ While Facebook Messenger has promised but not yet moved to end-to-end encryption, Facebook Messenger created similar barriers to monitoring by allowing users to turn on encryption using its “Secret Conversations” option.³⁶ The report concluded: “Without the ability to explicitly flag images or videos as CSAM, companies limit their capacity to remove offending content quickly.”³⁷

196. This is particularly problematic in a context in which Meta has repeatedly assured its users and the public that all is well with regard to CSEC on its platforms. Given that children, especially vulnerable children most likely to be susceptible to grooming, typically lack the insight and confidence to identify and report inappropriate activity, telling them that there is no issue, rather than encouraging them to be vigilant, arming them to recognize misconduct, and assuring them that Meta wants to hear and will act on their concerns, chills reporting and makes it less likely that Meta can detect and address unlawful and unwelcome content and conduct.

³³ Canadian Centre for Child Protection, *Reviewing Child Sexual Abuse Material Reporting Functions on Popular Platforms* (2020), https://content.c3p.ca/pdfs/C3P_ReviewingCSAMMaterialReporting_en.pdf, at 10.

³⁴ *Id.* at 12.

³⁵ *Id.* at 17.

³⁶ *Id.*

³⁷ *Id.* at 26.

197. Internal Meta documents provided to the Journal indicate that, in order to cut costs, Meta actually made it more difficult for users to report content in order to discourage such reports.³⁸ The internal documents reference an effort to “add thoughtful friction that reduces the number of spurious user reports we receive,” but that same document acknowledges that “we may have moved the needle too far on adding friction,” thereby preventing useful and important reports from making their way to individuals tasked with reviewing such content.³⁹

198. Meta knew that user reports undercounted harmful content and experiences on its platforms, but nonetheless made it harder, not easier to report and act on this information. In September 2021, Meta consultants, including Arturo Bejar (whose experiences are discussed below), commissioned a study to evaluate user experiences on its platforms, called “BEEF” (an acronym meaning Bad Experiences and Encounters Framework). The purpose of the study was to “develop a holistic, consistent picture of user bad experiences on Instagram” and create a metric that allow Meta to “track our progress” in eliminating those experiences. The BEEF survey demonstrated that 51% of Instagram users had a bad or harmful experience on Instagram in the last 7 days. Meta knew that only 1% of users reported the content or communication, and that only 2% of that user reported content is taken down. In other words, for every 10,000 users who have bad experiences on the platform, 100 report and 2 get help. Consistent with its public statements, a company that is both driven by data and genuinely committed to a product that is safe and good for children would have ensured that harmful content and interactions were reported and addressed.

199. Meta not only made user reports more difficult but, as described below, also dramatically cut its human investigators and content moderators, who are vital front lines in

³⁸ <https://www.wsj.com/tech/Instagram-facebook-teens-harassment-safety-5d991be1> (last visited Nov. 30, 2023).

³⁹ <https://s.wsj.net/public/resources/documents/metacostscontrols.pdf> (last visited Nov. 30, 2023).

detecting CSAM and trafficking. Also undermining Meta’s efforts to ensure that it would not allow CSAM on its platforms, Meta’s rules were narrowly written so as to prohibit “only unambiguously vile material.” Unacceptable, first-step grooming behaviors, including “adults . . . flooding the comments section on a teenager’s posts with kiss emojis or” sending invites to “see more” in private Facebook Messenger groups were not prohibited.⁴⁰

200. Hany Farid, who developed PhotoDNA, which is used by Meta and other technology companies to identify CSAM, was quoted to note that Meta could be taking more steps to develop tools to “flag suspicious words and phrases . . . – including coded language around grooming . . . This is, fundamentally, not a technological problem, but one of corporate priorities.”⁴¹

201. Arturo Bejar, who was Meta's Director of Engineering from 2009 to 2015 and then a consultant from 2019 to 2021 and recently testified before Congress concerning Meta’s problematic corporate conduct, confirmed that Meta’s internal and external reports touting the results of their automated enforcement relied on “sleights of hand.” The automated enforcement systems only captured information that Meta actually removed, not the majority of problematic content appearing on the platforms. Moreover, user reports and internal investigators are critical to ensure that Meta’s automated detection are effective in incorporating and addressing the current shape and pattern of problematic activity on its platforms.

202. Bejar’s testimony made explicitly clear that Meta’s public representations that CSAM and illicit content are rarely present on its platforms were unfounded and untruthful.

203. The combination of narrow automated detection, which fail to detect and remove a huge volume of CSAM, and Meta’s algorithms, which amplify and connect users to unaddressed

⁴⁰ <https://www.wsj.com/tech/Instagram-facebook-teens-harassment-safety-5d991be1> (last visited Nov. 30, 2023).

⁴¹ Guardian at 9.

CSAM, combine to produce the explosion of CSEC found by the Attorney General and reported by other observers.

IX. META HAS LONG BEEN ON NOTICE OF, BUT FAILED TO ADDRESS, COMMERCIAL SEXUAL EXPLOITATION AND TRAFFICKING OF CHILDREN THROUGH SOCIAL MEDIA AND ITS PLATFORMS SPECIFICALLY

204. Facebook has repeatedly and forcefully—but deceptively—misrepresented that it does not permit human trafficking, CSAM or CSEC on its platforms. As noted below, Defendant Zuckerberg personally stated in a Facebook post: “It’s very important to me that everything we build is safe and good for kids.” Responding to an article in the Guardian, “How Facebook and Instagram became marketplaces for child sex trafficking,” which was published in 2023, a Meta spokesperson promised: “The exploitation of children is a horrific crime – we don’t allow it and we work aggressively to fight it on and off our platforms.”⁴²

205. Meta has long been aware of the risk—and fact—of child grooming and trafficking, CSAM, and other sexual exploitation and exposure of children on its platforms, through the explosive news articles (and others) described in this Complaint and through publicly available reports. As laid out below, Meta’s own internal documents reveal that the company recognized that these issues were neither theoretical nor anecdotal—that trafficking, CSAM, CSEC, and other illicit activities were serious, ongoing, and systemic problems not adequately addressed by current measures. But Meta nonetheless chose not to take steps to prevent further harm to its young users or to honestly disclose to them or the public what it *knew* what happening to them.

206. Meta knew that adults soliciting minors was a problem on the platform, and was willing to treat it as an urgent problem when it had to. One internal document shows Meta scrambling in 2020 to address an Apple executive whose 12-year old was solicited on the platform,

⁴² <https://www.facebook.com/zuck/posts/10113961365418581> (last visited Nov. 30, 2023).

noting “this is the kind of thing that pisses Apple off to the extent of threatening to remove us from the App Store,” asking whether there was a timeline for when “we’ll stop adults from messaging minors on IG Direct,” and noting that if they did not address other accounts with “sugardaddy,” “they will reply with 100 more accounts if we’re not able to take these down.”

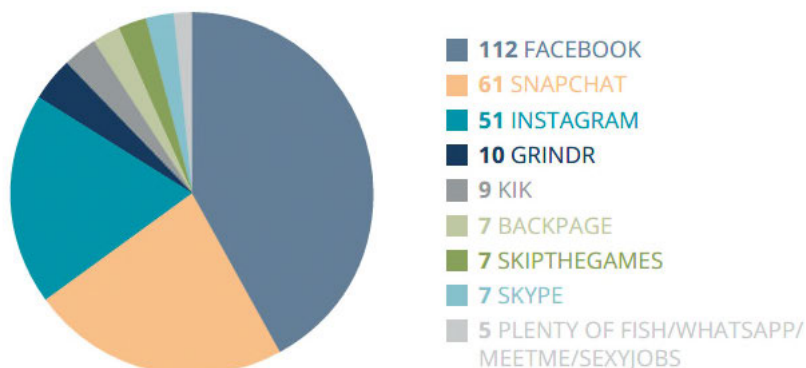
207. A 2018 Polaris study confirms that human trafficking often originates on social media “with the trafficker and potential victim building a relationship,” through interactions unique to social media, such as “commenting on potential victims’ photos and sending direct messages.” Indeed, 26% of the participants in a survey of trafficking survivors confirmed that “their trafficker exploited them via their own personal social media accounts.”⁴³

208. The 2022 Federal Human Trafficking Report (“FHTR”) identified the internet as the most common pipeline for recruiting victims of trafficking, which has been the case *for 23 years*. Focusing on the specific platforms most frequently used to engage in sex trafficking, the FHTR reported that Facebook, both on its own and together with Instagram, were the platforms most frequently used for recruiting victims of human trafficking between 2019 and 2022, nearly double all of the other platforms put together.⁴⁴

⁴³ Polaris Report at 22.

⁴⁴ Human Trafficking Institute 2022, *Federal Human Trafficking Report*, pp. 4-33. Available from: <https://traffickinginstitute.org/wp-content/uploads/2022/09/2021-Federal-Human-Trafficking-Report-WEB-1.pdf>. [13 March 2023].

TOP 10 PLATFORMS USED IN RECRUITMENT OF VICTIMS 2019-2022



209. Though Facebook says it doesn't tolerate child exploitation on its platform, another analysis by the Tech Transparency Project ("TTP") of federal criminal cases "show[s] pedophiles have inundated the social network for years."⁴⁵ In its review of U.S. Department of Justice press releases between January 2013 to December 2019, TTP found 366 federal criminal cases that involved the use of Facebook for child exploitation, "including distributing sexual abuse images, recruiting children and sex trafficking."⁴⁶ Child exploitation cases involving Facebook grew from 10 per quarter in 2013 to 23 per quarter in 2019.⁴⁷ Nearly half of those cases involved CSAM, and 41% communicating with or grooming children.⁴⁸ These statistics "represent the tip of the iceberg of a far larger problem" and do not count all federal investigations or violations of state law.⁴⁹

210. But "[o]nly 9% of the cases were initiated because Facebook or the National Center for Missing and Exploited Children (which receives tips of potential child trafficking or abuse, including from social media platforms and conveys them to law enforcement) reported them to

⁴⁵ https://techtransparencyproject.cdn.prismic.io/techtransparencyproject/38e39f6d-001d-466b-b9c0-b96eb5e57de9_Facebook-Child-Exploitation.pdf at 2.

⁴⁶ *Id.* at 2.

⁴⁷ *Id.*

⁴⁸ *Id.* at 7.

⁴⁹ *Id.* at 3.

authorities, raising questions about the effectiveness of Facebook’s monitoring of criminal activity targeting children.”⁵⁰ Rather than cooperation from Meta, “officials said they relied on information from the public, leads from other investigations or sting operations to identify suspects.”⁵¹ A NCMEC spokesperson also offered that “if social media companies are not reporting child sex trafficking, it allows this crime to thrive online. Reporting trafficking . . . is crucial for rescuing victims and punishing offenders.”⁵²

211. High profile cases filed in 2018 against Facebook and Instagram in Texas, which ultimately landed in the state and U.S. supreme courts, described in detail the manner in which young teenage girls were trafficked on Facebook. One case involved a 15-year old girl friended by a Facebook user whose profile featured photographs of “‘scantily-clad young women in sexual positions’ with money stuffed in their mouths, as well as ‘other deeply troubling content.’”⁵³ The adult user contacted the girl through Facebook’s “messaging system,” telling her that she was “‘pretty enough to be a model’ and promising to help her pursue a modeling career.”⁵⁴ The girl ultimately agreed to meet him and was forced instead into commercial sex, raped, and beaten.⁵⁵ A second case consolidated in the same appeal was that of a 14-year old girl who was contacted through Instagram by an adult male who trafficked her for prostitution on Instagram.⁵⁶ Even after she was rescued, the traffickers used the girl’s profile to lure other young girls.⁵⁷ The girl’s mother reported the activities to Facebook, which never responded.⁵⁸ The third matter related to another 14-year old girl, who accurately identified her age on her Instagram account, which was not

⁵⁰ *Id.*.

⁵¹ *Id.* at 2.

⁵² Guardian at 7-8.

⁵³ *In re Facebook, Inc.*, 625 S.W.3d 80, 84 (Tex. 2021).

⁵⁴ *Id.*

⁵⁵ *Id.*.

⁵⁶ *Id.*

⁵⁷ *Id.*

⁵⁸ *Id.*.

required to be linked to a parent account, and was friended by an adult stranger who persuaded her to meet him.⁵⁹ She, too, was offered for sale and raped repeatedly by men who responded to online ads.⁶⁰ Each of these cases demonstrated systemic failures by Meta that still exist and originated with conduct like that documented in this Complaint.

X. META PROFITS FROM THE ALL-TOO-COMMON DISTRIBUTION AND SOLICITATION OF CSAM

212. Distribution of CSAM and non-consensual intimate imagery (or “NCII”) is a known key aspect of human trafficking. Predators insinuate their way into private groups by “friending” one member with a public profile and then work their way through that person’s friends looking for new targets to victimize. Predators seek to gain the trust of their newfound “friends” and convince them to send naked pictures. Children and adolescents are especially vulnerable to these advances. Human traffickers are known to use threats to distribute such material as a method of controlling their victims, consistent with cases described above.⁶¹

213. Dr. Lisa Strohman is a clinical psychologist and founder of the Digital Citizen Academy, one of the first organizations to address issues of technology addiction and overuse. She specializes in treating children who have been victimized by social media, and, in her experience, illicit material on social media websites is widespread. Dr. Strohman estimates 80-85% of her teenage patients have been exposed to some form of CSAM, grooming, or sextortion on social media platforms, with Instagram being the worst of the platforms.

214. Absent appropriate vigilance, such material is freely shared among users, including individuals under the age of 18. A 2023 Stanford report concerning exchange of “self-generated child sexual abuse material” found that “Instagram is currently the most important platform for

⁵⁹ *Id.* at 84-85.

⁶⁰ *Id.*

⁶¹ Polaris Report at 25.

these networks” and that “Instagram’s recommendation algorithms are a key reason for the platform’s effectiveness in advertising” self-generated child sexual abuse material.⁶² That same report noted that, in stark contrast to Instagram, the researchers did not find similar material on “OnlyFans,” an adult-oriented website featuring sexual content that “has strict age verification and rules against the use of its platform by minors.”⁶³

215. Unlike OnlyFans, Meta has chosen to employ *no* age verification technology when users sign up for accounts. Its platforms lack this basic infrastructure even though Meta has been aware for years that children, including children under the age of 13, register for its platform by lying about their real age. A May 2018 presentation to Meta’s Audit Committee confirms this fact: “[W]hile user-provided data indicates a decline in usage among younger users, this age data is unavailable because a disproportionate number of our younger users register with an inaccurate age.”

216. Two years later, a January 2020 presentation entitled “Succeeding in US Messaging” demonstrated both the depth of Meta’s knowledge regarding usage of Messenger by children and its ambitions to leverage that usage to further entice younger generations to use its products. One of Meta’s “End Game” goals was to “[b]ecome the primary kid messaging app in the US by 2022.” The document confirmed that “[i]n the US,” “Messenger is popular . . . with Kids (13% primacy, US 6-10).” Meta’s knowledge that its platforms were used by and “popular with” children as young as 6-years old makes its failures to protect minors against CSAM and solicitation all the more egregious.

217. By August 2020, so many potentially underage individuals opened accounts on Meta’s platforms that Meta established an “Underage Enforcement War Room” in order to keep

⁶² Stanford SG-CSAM Report.

⁶³ Stanford SG-CSAM Report.

up with a growing backlog of more than 1,160,000 accounts requiring review. The backlog existed because of “lack of monitoring in this space due to it being a lower priority area,” as well as personnel changes during the COVID lockdown. As noted below, the backlog only increased in following years.

218. Automated procedures reduced the “enforcement backlog” to “only” 700,000 potentially under-13 users. However, a November 2020 internal presentation cited “poor cross app enforcement” with respect to identifying underage accounts and noted that Meta’s “Age models” were “Not very accurate, not globally resilient and not globally operational” and also had “uniquely poor efficacy on IG:”

Age Assurance/Appropriateness State of Play	
Vulnerabilities	Opportunities
<ul style="list-style-type: none"> • U13 enforcement: lack classifier, no way to detect and remove at scale; U13 700k enforcement backlog and poor cross app enforcement; deprioritized enforcement due to concern re: legal liabilities • U18 enforcement: age gating relies on either stated age or weak age models; lack checkpoint • Age verification: pushed to device, slow advancement, dependent on x-industry cooperation • Age models: not very accurate, not globally resilient and not globally operational; weak models negatively impact other integrity measures, including safety classifiers; uniquely poor efficacy on IG • Defaults, gating, rules: inconsistent, need wholesale review • Minors in-product education: minimal • Content: inappropriate/harmful content and experiences for minors 	<ul style="list-style-type: none"> • Age appropriate experiences • Unlock interoperable aging up (e.g., MKids to tween experience to IG & FB) • Device based age verification

219. In January 2021, Meta acknowledged (internally) the danger of relying on a user’s stated age, noting that “[s]tated age only identifies 47% of minors on [Instagram]. The majority of minors lie when stating their age.” Moreover, “99% of disabled groomers do not state their age .

. . The lack of stated age is strongly correlated to IIC [inappropriate interactions with children] violators.”

220. The “backlog” of enforcement actions against under 13 users only got worse over the years. By March 2021, internal Meta documents reported that at least *2.5 million* accounts flagged as potential underage users were awaiting review. The backlog existed, in part, due to Meta’s refusal to appropriate sufficient resources (as well as its failure to institute appropriate age verification measures at sign up). A presentation entitled “Integrity Update” reported that the group was awaiting approval to “increase our supply of human reviewers to burn down the backlog.” That the backlog extended to 2.5 million accounts and approval for more resources was not forthcoming serves as further proof that Meta valued profits over illegal use of its platform by underage children.

221. The lack of sufficient age verification mechanisms is even more troubling given the findings of the Stanford Cyber Policy Center that distribution of self-generated CSAM proliferates on Instagram. The Stanford report found “405 accounts advertising the sale of self-generated CSAM on Instagram” and numerous “probable content buyers” who operated using their own names. After these accounts were reported to authorities, “hundreds of new SG-CSAM accounts were created, recreated or activated on” Instagram, and “linked to the network” by hashtags or other content.

222. Indeed, according to the Stanford report, merely “limiting” accounts trafficking in self-generated CSAM did little to stem the tide of illicit material on social media platforms. Following such actions, “sellers either switch to a backup (often previously noted on their main

profile) or make a new account, which is then promoted in Instagram Stories . . . by other accounts in the network to help it regain lost followers.”⁶⁴

223. Dr. Strohman has had similar experiences with reporting illicit activity to social media operators, including Meta. In her experience, most content is not taken down after being reported and Dr. Strohman herself has reported numerous users or posts to Meta and never received any response. She has also reported groups that list “children” who are to be sold. If the sites direct users to a WhatsApp number, in her experience, that is a clear red flag of illicit activity.

224. The State, upon information and belief, understands that Meta profits from the human trafficking, CSAM and CSEC activity appearing on its platforms. One undated internal document openly acknowledged that Meta faced “reputational risk” from “accrued ad revenues associated with HT [Human Trafficking]”:

IV - RECOMMENDED ACTION ITEMS				
<p>***If you see your name below, it means that we believe that you're probably the best person/team POC that could help FB move forward on tackling sex trafficking in respective areas post this investigation. Please help loop in additional people or redirect if we've misidentified you (nonetheless thank you for reading!). Master Task: T53526763</p>				
ACTION ITEM	RISK/ IMPACT	AI TYPE	OWNER	TASK
Clarify Ads Account policy why/when payment methods related to ad credit accounts are allowed with no names/ info	Prevent traffickers from using multiple CC with nameless payment methods linked to the ad credit accounts & close any policy gap	Policy	ad-policy	T53526846
Clarify policies how FB currently handles accrued ad revenues associated with HT	Prevent reputational risk for the company (not to profit from ads spent for HT)	Policy	[REDACTED] / Antigone Davis / [REDACTED]	T53526954

225. Another document dated in or about 2020 acknowledged that Meta received profits attributable to human trafficking: “We know we don’t want to accept/profit from human

⁶⁴ Stanford SG-CSAM Report at 6.

exploitation. How do we want to calculate these numbers and what do we want to do with this money (i.e. from ad spend)?”

226. But Meta also profits in other ways from human trafficking content and distribution/solicitation of CSAM or CSEC on its platforms. The predators who distribute or solicit this content are users nonetheless, and thus potential targets for advertisers. As Meta warned in its financial statements, if these users were to leave Meta’s platforms or were to become less engaged, that could result in reduced revenue for the company from advertising sales. Moreover, users can purchase “stars” or other “gifts” from Meta that can be awarded to other users. Nothing prevents such “gifts” from being awarded to or from sexual predators. Thus, while Meta publicly positions itself as vigilant against this type of content, the findings noted above and discussed herein indicate that Meta has placed profits before safety in its enforcement efforts.

XI. CHILDREN IN NEW MEXICO HAVE FALLEN VICTIM TO DISTRIBUTION AND SOLICITATION OF CSAM OR CSEC ON META’S PLATFORMS

227. New Mexico’s children have been trafficked and sexually exploited through Meta’s platforms. The New Mexico Human Trafficking Task Force has observed that “[t]raffickers hit social networks to recruit underage girls to engage in commercial sex” and “searched Facebook for attractive young girls, and sent them messages telling them that they were pretty and asking if they would like to make some money.”⁶⁵

228. Cases dating back more than a decade and continuing through this year, involve children in New Mexico who were contacted, groomed, and lured into sending explicit photographs or meeting predators through Facebook or Instagram. These include a 17-year old girl

⁶⁵ New Mexico Human Trafficking Task Force, *Human Trafficking in New Mexico, Presentation For: Courts and Corrections Justice Committee*, <https://www.nmlegis.gov/handouts/CCJ%20101617%20Item%205%20Human%20Trafficking%20in%20New%20Mexico.pdf>

who was contacted after posting a notification on Facebook seeking a job (consistent with conduct the Attorney General’s investigation again encountered). After false promises, the trafficker pressed her into commercial sex and tried to sell her as a “sex slave.”⁶⁶ Police pulled over an adult male with six children in his car, including a 13-year old girl he met through social media, communicated with through Snapchat and Instagram, and raped repeatedly.⁶⁷ Another child predator recruited young girls by starting conversations via Facebook and then migrating the conversations to text messages and Snapchat. He reached more than 100 victims, including in New Mexico, pretending to be an 18- or 19-year old high school student and then meeting them at malls or picking them up at school.⁶⁸ Tragically, there are other stories like these.

229. A complaint pending in multidistrict litigation against Meta and other social media platforms, tells the story of one New Mexico teen, L.K, from Rio Rancho, New Mexico.⁶⁹ According to the complaint, L.K. began using social media, including Instagram, when she was around 9- or 10-years old. Her use of social media became compulsive, late into the night, and her performance at school began to suffer. Her parents frequently took away her phone. Over time, L.K. developed depression, anxiety, eating disorders, and isolation. During high school, L.K. was induced to send sexually explicit pictures to two individuals through Snapchat. They used threats of distributing the photographs to force L.K. to engage in sex. Despondent, L.K. died by suicide at age 17.

230. Human trafficking has had a particularly profound impact on Native communities in New Mexico. Native Americans are estimated to represent nearly one-fourth of trafficking

⁶⁶ <https://www.koat.com/article/police-man-tried-to-sell-teen-into-sex-slavery/5037710> (last visited Dec. 1, 2023).

⁶⁷ Germania Rodriguez Poleo, *Shocking moment ‘child rapist’ is found with six children in his car and his pants unbuttoned after being pulled over while drink driving – before saying the kids are ‘just my friends’*, DailyMail.com (June 28, 2023), <https://www.dailymail.co.uk/news/article-12242713/Moment-child-rapist-six-children-car-pulled-over.html>.

⁶⁸ Anthony Cotton, “Suspect Trolls Facebook for girls,” Denver Post, June 26, 2015.

⁶⁹ *Youngers v. Meta Platforms Inc., et al.*, 1:22-cv-00608-KWR-LF, filed 8/15/22.

victims, more than double their share of the state's population.⁷⁰ The Coalition to Stop Violence Against Native Women lists social media as one of the venues through which sex trafficking occurs, including through Facebook.⁷¹ As far back as 2018, a Navajo Nation leader acknowledged the role played by social media: "Our Navajo children are being picked up through social media and trafficked at truck stops or other areas across the United States."⁷²

231. One publicly reported case illustrates that trajectory. Eva was given an iPhone for her 12th birthday to help keep her safe when she was home alone. In 2015, when she was just 13, a man messaged her on Facebook claiming to know her from school. He complimented her appearance and eventually persuaded her to send him sexually explicit photographs and then to meet him. He recorded Eva performing oral sex and having intercourse with him and, as with L.K., threatened to publish the photos and videos on Facebook if she said anything. He brought in other men who forced her to have sex with them, too. Eva ultimately made it to a safe house and broke free of her trafficker. In 2019, while in the 10th grade, she confided in a reporter: "I want to make it not real. But I was living there. And sometimes, I'm still living there."⁷³

232. The trauma experienced by children who are subjected to commercial sexual activity is self-evident and inarguable. Studies have examined the profound, life-long impact that publication of CSAM and other sexual images alone has on its victims.⁷⁴

⁷⁰ *Stolen and Erased / Navajo Girl Exploited, Sex Trafficked for Years*, Rio Grande SUN, 01/02/2020.

⁷¹ https://cspanw.org/wp-content/uploads/2019/03/TLS_STTC_Flyer-F.pdf.

⁷² 23rd Navajo Nation Council, "Navajo Nation branches stand in solidarity to end human trafficking," Jan. 26, 2018, available at: https://www.navajonationcouncil.org/wp-content/uploads/2020/09/Navajo_Nation_branches_stand_in_solidarity_to_end_human_trafficking.pdf (last visited Dec. 2, 2023).

⁷³ <https://searchlightnm.org/stolen-and-erased/> (last visited Dec. 1, 2023).

⁷⁴ Cites in this paragraph + ECPAT International (2018), *Towards a Global Indicator on Unidentified Victims in Child Sexual Exploitation Material*, citing Nyman, A (2008) "Abused Online," Svedin and Back (1996), "Children who don't speak out"; Loof (2005), "Global Issues and Regional Co-operation in Fighting Child Exploitation"; Brennan and Phippen (2018), "Youth-Involved Sexual Imagery;" Cooper (2012), "The impact on children who have been victims of child pornography;" Gewirtz-Meydan, Ateret, Yael Lahav, Wendy Walsh, and David Finkelhor. "Psychopathology among Adult Survivors of Child Pornography." *Child Abuse & Neglect* 98 (December 1, 2019): 104189. <https://doi.org/10.1016/j.chiabu.2019.104189> (last visited Dec. 1, 2023).

233. Victims of CSAM suffer the initial crime, are retraumatized each time the CSAM is republished, and live in fear that they will be recognized from the images and often in silence and shame because of threats or fear that reporting will cause images of them to be further distributed. According to one CSAM survivor: “It’s hard to describe what it feels like to know that at any given moment someone somewhere is looking at images of me as a child being sexually abused and getting sick gratification from it. It’s like I’m being abused over and over and over again.”⁷⁵ In the words of another:

I was a victim, but now a survivor of child sexual abuse and sexual abuse imagery that has traveled the world via the internet. Abuse that began when I was an infant until I was 4 years old. I was abused by my ex-father and one of his friends. The imagery of me is still being shared on the internet. To date, I have received over 35,000 notifications from the U.S. DOJ regarding those who are in receipt of my abusive images. It’s not fair. No child should have to endure the pain, the hardships, the loss of innocence or a normal life because of the hands of an abuser and those who take pleasure from the suffering of children.”⁷⁶

234. In other studies, CSAM survivors describe feeling ashamed, guilty, humiliated, and “haunted,” worried not only about who would see the images but that people would think they were willing participants.⁷⁷ A survey of survivors conducted by the Canadian Centre for Child Protection in 2017 found that “[n]early 70% of respondents indicated that they worry constantly about being recognized by someone who has seen images of their abuse.” Fifty-six percent of those who participated in the survey were between the ages of 0 to 4 when the abuse started, and two-thirds reported being threatened with physical harm if they didn’t comply with abuser’s wishes or if they told someone about the abuse.⁷⁸

⁷⁵ <https://www.protectchildren.ca/en/> (last visited Dec. 1, 2023).

⁷⁶ https://content.c3p.ca/pdfs/C3P_ReviewingCSAMMaterialReporting_en.pdf, at 26 (last visited Dec. 1, 2023).

⁷⁷ Gewirtz-Meydan, Ateret, Wendy Walsh, Janis Wolak, and David Finkelhor. “The Complex Experience of Child Pornography Survivors.” *Child Abuse & Neglect* 80 (June 1, 2018): 238–48.

⁷⁸ Canadian Centre for Child Protection, *Survivors’ Survey: Executive Summary 2017* (2017), https://protectchildren.ca/pdfs/C3P_SurvivorsSurveyExecutiveSummary2017_en.pdf (last visited Dec. 1, 2023).

235. A recent Washington Post article chronicled the stories of teenagers subject to sextortion on social media online, which sounded similar themes. According to the executive director of the Exploited Children Division at NCMEC, the number of sextortion cases “has exploded in the past couple of years.” The article reports that the Center received more than 10,000 tips of financial sextortion of minors, primarily boys, in 2022, and received more than 12,500 reports through July of this year. Sextortion has devastating effects on its underage victims. The Washington Post report cited at least a dozen suicides by teenage boys in 2022 due to sextortion. Each of the victims interviewed in the article met their scammer over the internet on social media platforms.⁷⁹

236. While Meta promised to safeguard the health and safety of children on its platforms (and to keep the youngest users offline), at every turn, it made decisions that put its own profits ahead of their well-being. Too many children in New Mexico, from L.K., to Eva, to the fictional profiles whose experiences online stood in for the experiences of real New Mexico children, have been casualties of Meta’s choices. Despite the reports, criminal cases, and news articles described above, including evidence that Meta’s platforms were and continued to be used to engage in human trafficking, Meta failed to take adequate measures to address and disclose human trafficking on its platforms.

XII. META WAS, AND IS, AWARE THAT ITS PLATFORMS ARE BEING USED TO TARGET, GROOM, SEXUALLY EXPLOIT AND TRAFFIC CHILDREN

237. Internal documents demonstrate that Meta has been aware for years that underage users regularly use its platforms, including Facebook and Instagram, and that both Facebook and Instagram were being used to target, groom, sexually exploit and traffic children. It further was

⁷⁹ Chris Moody, “‘IDK what to do’: Thousands of teen boys are being extorted in sexting scams,” Washington Post (Oct. 2, 2023).

aware that certain features of its platforms, including the algorithm, actively encouraged this conduct and recommended it to children and to users with nefarious motives. Yet despite this knowledge, Meta failed to take steps to improve the design of its products and illicit activities on Facebook continue to flourish to this day.

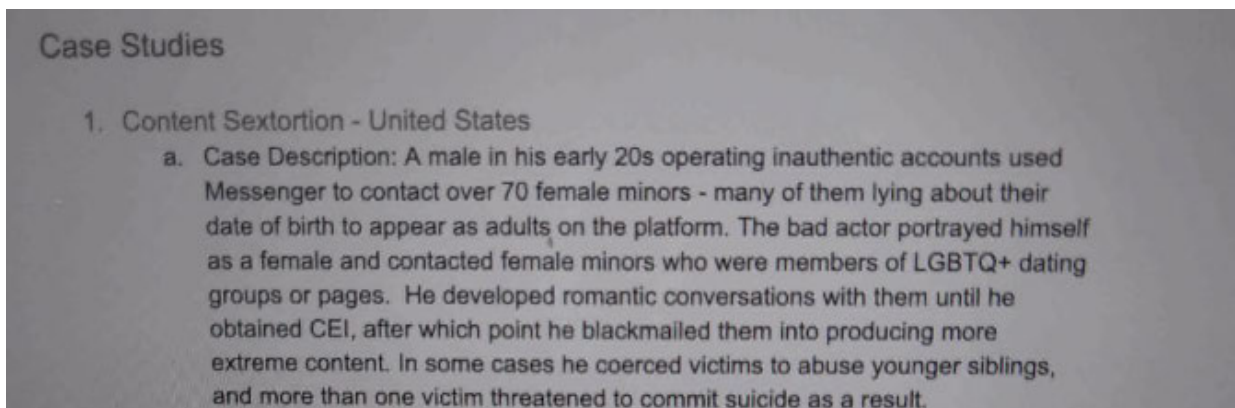
238. That Meta was aware of the presence of this disturbing content is significant. Testimony by whistleblower Frances Haugen confirmed that Meta was aware that “Facebook’s AI systems only catch a very tiny minority of offending content” and reliance “on computers and not humans” means that Meta will “never get more than 10 to 20%” of the illicit content on its platform.

239. One 2019 internal document describes the most frequent forms of human trafficking on Meta’s “FOAs” or family of applications. Though recognizing that “every Human Exploitation stage (recruitment, coordination, exploitation) is represented on our platform,” the company determined that recruiting and exploiting (or advertising) victims for profit were the most common. Meta noted that it had observed traffickers using romance to build trust and rapport with potential victims; using lucrative employment opportunities on FB [Facebook] Groups of Ads from Pages to entice potential victims, “[t]argeting the most vulnerable sector of the population usually because of their precarious economic situation;” and using Messenger “to coordinate trafficking activities”—all activities still observed in the State’s investigation.

240. In the same document, Meta also recognized that Facebook Dating “will be used by Human Exploitation actors, particularly around sex trafficking” to “target new potential victims” and “connect those victims with potential buyers.” In the same vein, Meta also recognized that its platforms would be used to solicit intimate images from children, starting with

“grooming/seduction” and then often moving to sextortion, forcing victims “to produce sexual content” on an ongoing basis.

241. Meta offered a case study of this form of “content sextortion” in the United States:



242. Public reports establish a long history of Meta recognizing the existence of harmful content on its platforms and entrenched, institutional resistance to taking actions to remediate the problems. As early as 2017, an internal email describes executive opposition to scanning the Meta Messenger service for “harmful content” because it would place the service “at a competitive disadvantage vs other apps who might offer more privacy.”

243. Meta does not review content before it is uploaded. Rather, as noted above, it relies on automated detection tools and user reports to identify objectionable material, which is then reviewed by hired “content moderators,” who are often subcontractors. A 2017 article in The Guardian confirmed that most of Meta’s “content moderation” (i.e. viewing posts and messages to determine whether the content is acceptable or illegal) is performed by subcontractors, who “often fe[lt] overwhelmed by the number of posts they ha[d] to review” and Facebook’s numerous, confusing policies.⁸⁰ Content moderation, in and of itself, is a job that has been linked to emotional trauma, including PTSD, from repeatedly viewing objectionable content and from the strict

⁸⁰ <https://www.theguardian.com/news/2017/may/21/facebook-moderators-quick-guide-job-challenges> (last visited Dec. 1, 2023).

demands placed on them to maintain certain review metrics.⁸¹ The Guardian’s reporting indicated that Meta’s guidance often permitted violent content to remain on the site. One slide provided that the statement “To snap a bitch’s neck, make sure to apply all your pressure to the middle of her throat” was acceptable and could remain on the platform.⁸²

244. The Guardian also spoke with a half-dozen sub-contracted moderators responsible for identifying child sex trafficking for Meta. As reported by the Guardian, these moderators expressed concerns that Meta failed to act on even the plainest indications of trafficking or child sex exploitation, let alone those in which traffickers made an effort to mask their conversations with code words or emojis, like those described above. One moderator, who spoke under a pseudonym reported that “she and her team struggled to keep pace with the huge backlog of cases. She says that she saw cases of adults grooming children and then making plans to meet them for sex, as well as discussions about payment in exchange for sex,” which is consistent with the State’s own findings.⁸³ She said that reports to Meta would linger for months, when she received an email saying that the case was closed because “nobody’s taken action on it . . . [I]t felt like nobody would pay attention to these horrible things.”⁸⁴ Another said, “On one post I reviewed, there was a picture of this girl that looked about 12, wearing the smallest lingerie you could imagine. . . . It listed prices for different things explicitly, like, a blowjob is this much. It was obvious that it was trafficking.” Yet her supervisor told the moderator that no further action was taken after the content was escalated.⁸⁵

⁸¹ <https://www.theguardian.com/technology/2017/may/04/facebook-content-moderators-ptsd-psychological-dangers>.(last visited Dec. 1, 2023).

⁸² <https://www.theguardian.com/news/2017/may/21/revealed-facebook-internal-rulebook-sex-terrorism-violence> (Dec. 1, 2023).

⁸³ Guardian at 9.

⁸⁴ *Id.*

⁸⁵ Guardian at 10.

245. According to documents produced by Meta whistleblower Frances Haugen, Meta's review of potentially violating content was also limited based on users' "expectations of privacy." Thus, Messenger groups with less than 32 people were to be "treated with a full expectation of privacy," presumably limiting the review of content that violated Meta's policies and the law within these groups.

246. These documents also demonstrate that Meta knew that trafficking activity also migrated and was carried out across its platforms. For example, one highlighted "investigation revealed that the final arrangements for the recruitment and the coordination of human exploitation activities occurred in WhatsApp."

247. Further illustrating both its recognition of a common pattern of conduct and its unwillingness to set its own lines against grooming, trafficking, and CSAM, Facebook in 2018 asked users how Facebook should handle "a private message in which an adult man asks a 14 year old girl for sexual pictures."⁸⁶

⁸⁶ <https://www.theguardian.com/technology/2018/mar/05/facebook-men-children-sexual-images> (last visited Dec. 1, 2023).

09:45

Survey

There are a wide range of topics and behaviors that appear on Facebook. In thinking about an ideal world where you could set Facebook's policies, how would you handle the following: a private message in which an adult man asks a 14 year old girl for sexual pictures.

☐ This content should be allowed on Facebook, and I would not mind seeing it

☐ This content should be allowed on Facebook, but I don't want to see it

☐ This content should not be allowed on Facebook, and no one should be able to see it

☐ I have no preference on this topic

Continue

09:46

Survey

When thinking about the rules for deciding whether a private message in which an adult man asks a 14 year old girl for sexual pictures should or should not be allowed on Facebook, ideally who do you think should be deciding the rules?

☐ Facebook decides the rules on its own

☐ Facebook decides the rules with advice from external experts

☐ External experts decide the rules and tell Facebook

☐ Facebook users decide the rules by voting and tell Facebook

☐ I have no preference

Continue

248. Guy Rosen, Vice President of Integrity at Facebook, put it most plainly in his 2018 response to an email noting that Facebook does not scan direct messages for violating conduct, such as grooming and solicitation. Rosen offered to provide Mosseri with a “shadowing session” to demonstrate the “kind of stuff” that investigators find. He described three tiers of inappropriate interactions with children: [REDACTED]

[REDACTED] this is the worst of the worst stuff that gives you nightmares, over which children have taken their own lives :(” Rosen advised that “we can’t

do much of anything on Tier 1 & 2 without proactive scanning [of messages]. Also, because [REDACTED]

Elsewhere, Meta acknowledged [REDACTED] of Tier 2 conduct on Instagram Direct, describing it as a “key risk facing the success of encryption.” Rosen concluded: “The meta way I think about it is that both FB & IG are discovery services, not just communication services; [REDACTED]

and warned that without action, “otherwise they are basically massive ‘victim discovery services.’”

249. Despite this inventory of the grave, life-threatening harms that Meta had observed on its platforms, all of which the Attorney General’s Office confirmed *still occur*, Meta still failed to take vigorous action and still continued to misrepresent the frequency and severity of CSEC on its platforms. Upon information and belief, Meta still does not proactively scan messages without the user reports that it acknowledged it rarely receives.

250. In June 2019, an executive with the National Center on Sexual Exploitation advised Meta in correspondence that “Sex trafficking on your platform is a common occurrence” and that the organization’s “president just raised this issue at the White House.”

251. An email from a reporter in October 2019 informed Meta that “Those who are familiar with sex trafficking of minors claim that they feel Instagram is not doing enough to ensure men and women cannot sell or buy sex from minors on the app.” The reporter questioned whether Meta placed restrictions on the hashtags “#sugarbaby” or “#sugardaddy.” Meta responded “There are no restrictions in place on the hashtags in question.” And, indeed, State investigators discovered the express use of the phrase “sugar baby” to solicit a minor.

252. In December 2019, Meta employees circulated a Sunday Times article that focused on Instagram’s algorithm recommending objectionable content. One individual in the article

“specifically calls for [Instagram] to stop recommending children’s accounts to anyone. They have a case study of a parent who set up an account to showcase her daughter at gymnastics and was horrified that this was then potentially promoted to other people.” The email further notes the existence of Instagram users under the age of 13 and “[t]he ease of which a stranger can direct message a child on Instagram and the lack of proactive protections in place.”

253. In July 2020, a document entitled “Child Safety – State of Play (7/20)” demonstrates further confirmation of the presence of illicit activity on Meta’s platforms, and Meta’s knowledge of the same. The list of “Immediate Product Vulnerabilities” acknowledges “livestreaming abuses,” negative impact on the ability to report content that disappears on Stories, and “interop” presumably referencing cross-platform conduct:

Immediate Product Vulnerabilities

- Rooms: livestreaming abuse
- Ephemerality: negative impact on ability to report
- Interop: discoverability and reachability

IAL - PROTECTED BY CONFIDENTIALITY AGREEMENT

MT-IG-AG-NM-000171160

- NCMEC reporting: missing context
- WhatsApp groups: link sharing on social platforms + apps on iOS and Android store enable CSAM discovery and sharing
- COVID review constraints
- Unconnected adults being able to find and message minors in Instagram Direct
- Implicit sexualisation of minors across multiple posts and comments on Instagram
- Sex trafficking and sexual solicitation networks on Instagram

These vulnerabilities remain today with respect to Facebook, Instagram, and WhatsApp. The document’s recommendations for “Areas of Improvement” includes topics such as “NCMEC

Reports,” “Minor Sexualization” and “Trafficking,” again demonstrating Meta’s awareness that it fell short in these critical areas.

254. The document further confirmed that safeguards appearing on Facebook were not present on Instagram, most notably the fact that “work is not planned to prevent FB and IG adults from initiating messages to IG minors,” even though Meta was working to prevent contact between adults and unconnected “minors on FB.” Indeed, the document reports that efforts to “Block reach of unconnected adults to IG minors” was work that “has not been prioritized and more resources are needed.” The author sharply criticizes the rationale for not “prioritizing” such work (curtailing the ability of adult relatives to reach minor family members) as “less than compelling” and confirms that Meta had sacrificed children’s safety for “a big growth bet.” The full excerpt appears below:

15	Not ready; work has not been prioritised and more resources are needed					
16	Aggregated content on IG surfaces	Build classifiers to detect and take action on problematic aggregated content on surfaces like Explore, hashtag pages, Reels				
17	Block reach of unconnected adults to IG minors	Currently, on MSGR, we do not deliver messages initiated by unconnected adults to minors. In IGD, we do not have this protection in place. For Interop, product has committed to preserving this rule in MSGR, but not for cross-network messages sent by FB adults to IGD minors. Nor will they protect IGD minors from receiving messages from unconnected IGD adults. In light of the fact that grooming on IG occurs most likely ~2x as much as FB, their main reason not to build this protection for IGD minors is less than compelling: we do not want to block FB parents and older relatives from being able to reach out to their IGD children/relatives. They claim this is one of the main value props for Interop and a big growth bet.				

255. Moreover, that same document confirms that Meta affirmatively decided not to take steps to prevent minors from accessing its platforms and failed to enforce any age restriction violations across its platforms. “In March it was decided that we would not build a classifier to detect minors under 13 years old . . . We need leadership support to push product to act on *knowledge we already have*, i.e. clear our backlogs, disable verified u13s, build an u18 checkpoint

to review minors who have been reported on 18+ products.” (emphasis added). As noted above, Meta’s “enforcement” of safety issues “is not consistent across all apps for a single known person. In addition, age models perform inconsistently across different markets.” The Meta employee confirmed these failures placed the company at risk of violating federal statutes concerning the provision of internet services to minors: “[O]ur basic COPPA compliance is at risk when product does not prioritize checkpointing and disabling u13s [children under 13].⁸⁷ We also do not yet cross-enforce against u13s discovered on IG who have hard-linked FB accounts.”

256. Age verification and preventing harm to minors was rendered even more difficult on Instagram because, as the “Child Safety – State of Play” document confirms, Instagram “only started to collect age data in December 2019 and we currently only have stated age for ~67% of IG users.” The document placed the blame directly on Instagram’s leadership, including Adam Mosseri, the head of Instagram, “We need leadership support to convince IG leadership to collect age for the remaining users for whom we have no stated age. We also would like leadership sign off to conduct an audit of special protections for minors and then bring in a trusted expert/group of experts to review our default settings and protections for minors, identify improvements . . . , and implement them.” On information and belief, this audit was never performed.

257. Further, the “Child Safety – State of Play” document confirmed that Instagram’s large safety holes were attributable to Mosseri’s failure to adopt safety recommendations:

⁸⁷ COPPA refers to the Children’s Online Privacy Protection Act, 15 U.S.C. § 6501 *et seq.*, which requires companies to notify and obtain consent from parents before collecting, using, or disclosing information from children under 13-years old.

5	Not ready; work has not been prioritised and more resources					
---	---	--	--	--	--	--

11

NTIAL - PROTECTED BY CONFIDENTIALITY AGREEMENT

MT-IG-AG-NM-0001

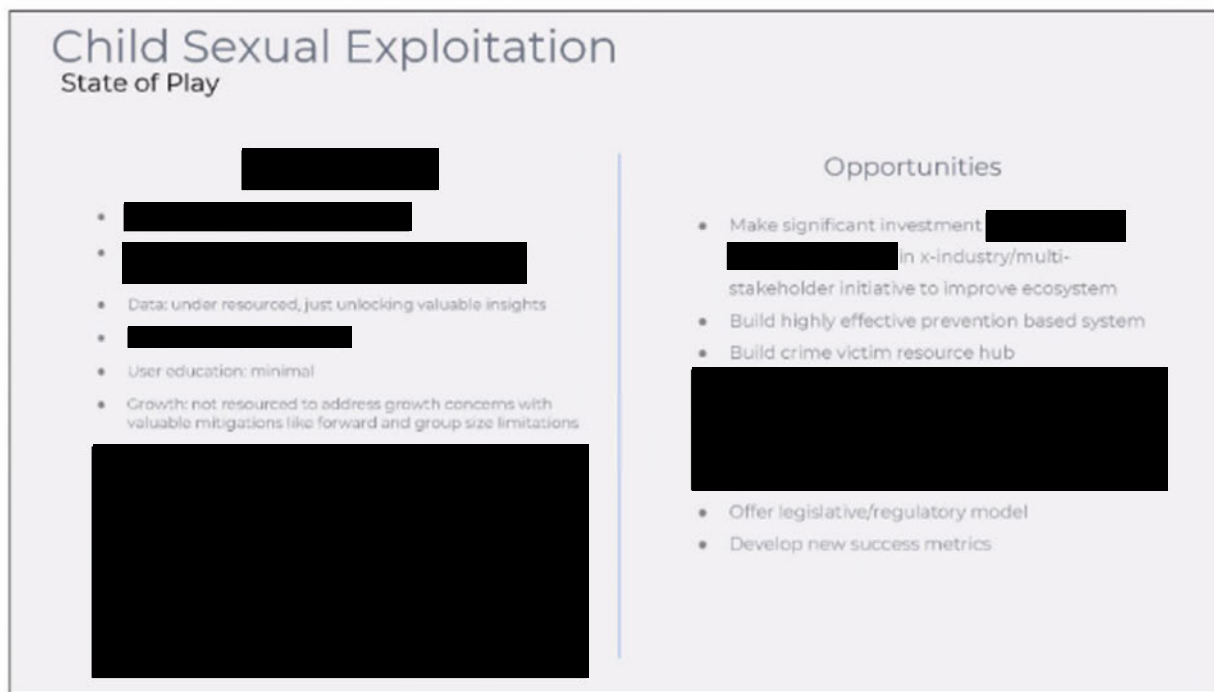
	are needed					
6	Age data on IG	Age models do not work as well on IG because IG only started to collect age data in December 2019. In addition, Adam Mosseri will not support upselling age collection for existing users because he sees it as intrusive data collection. Thus, we currently only have stated age for ~67% of IG users. We recommend engaging with IG leadership on collecting age for the remaining ~33% and prioritizing age model maturity on IG.				

258. Internal documents confirm Meta’s failure to prioritize reducing harm to children. In a July 2020 internal Meta chat, one employee asked “What specifically are we doing for child grooming (something I just heard about that is happening a lot on TikTok)?” He received a response—“somewhere between zero and negligible”—that underscored Meta’s lack of investment:

██████████ (7/09/2020 11:42:58 PDT):
 >What specifically are we doing for child grooming (something I just heard about that is happening a lot on TikTok)?

██████████ (7/09/2020 11:43:36 PDT):
 >Somewhere between zero and negligible. Child safety is an explicit non-goal this half. I'd argue we're making it worse with Interop, but's that's a can of worms.

259. Meta's employees continued to raise alarms regarding Instagram's inability to police inappropriate and illicit sexual activity targeted to children on its platform. A September 2020 internal email reporting results from "a one-off prevalence analysis" advised that "prevalence of 'sex talk' to minors is **38X** on IG Direct vs. Facebook Messenger, in US." The email urged Meta employees to implement further safeguards on Instagram accounts. Those recommendations were not implemented for a significant period of time though Meta knew that "inappropriate interactions with children aka child grooming" was "out of control" on Instagram. More than one-quarter (28%) of minors received message requests from adults they did not follow on Instagram. A November 2020 presentation also entitled "Child Safety: State of Play" (but separate from the chart described above) underscored the abundance of harmful content appearing on Meta's platforms. As one slide confirmed, Instagram employed "Minimal child safety protections," policies regarding "Minor sexualization" were described as "immature," and there was a "Minimal focus" on "Trafficking:"



260. Meta was also well aware that barriers existed to reporting content on Instagram. In a December 2020 document, one employee noted “[f]indings from a month long diary study.” The “Key Takeaways” included: “People often see content that shouldn’t be on Instagram, but they don’t always report it” and that “Instagram needs a unified reporting system.” The summary further noted that “People lack trust that reporting will be effective . . . This perception may result from past experiences with Instagram inaction on reports, and their doubt in the fairness of our content moderation process.” The employee also reported that “[t]he reporting flow can be too complex to use and understand, particularly for teens.”

261. In February 2021, while investigating an account reported by “a national news outlet” that it had investigated a report for potential child pornography “and immediately receiv[ing] messaging that it did not violate [Meta’s] Community Standards,” Meta employees “found a wide network of related accounts engaging in similar behavior, which shows us this is

not an isolated incident.” It is unclear what, if any, safeguards were implemented after this finding, but the State’s findings detailed above demonstrate that Meta’s efforts were unsuccessful.

262. Another presentation on Child Safety, this one dated March 2021, provides even further confirmation of the impacts of Meta’s historical reluctance to institute appropriate safeguards on Instagram. The presentation confirms that Meta’s “measurement, detection and U18 product safeguards are more mature on FB & MSGR than IG,” thus demonstrating that Meta failed to apply known protections across its platforms to Instagram. Under the heading “IG-specific challenges,” the slide confirms that inappropriate interactions with children, which the presentation refers to as “IIC, a.k.a. ‘grooming’” exists on Meta’s platforms. “Having been historically blocked from labelling message content (for privacy reasons), our IIC modes aren’t as mature.” The presentation further notes that Meta is “underinvested in minor sexualization on IG, notable on sexualized comments on content posted by minors. Not only is this a terrible experience for creators and bystanders, it’s also a vector for bad actors to identify and connect with one another.” In other words, Meta recognized that its platforms facilitated grooming, and that steps were required to address the harms that resulted.

263. Another 2021 document reflected Meta’s awareness that its recommendation algorithm, People You May Know, or PYMK, had a direct link to trafficking: In a string of comments, under the heading “IIC/Grooming,” a Facebook employee wrote: “in the past, PYMK contributed up to 75% of all inappropriate adult-minor contact,” prompting another employee to pose the question: “How on earth have we not just turned off PYMK between adults and children? . . . It’s really, really upsetting.”

264. In 2021, Meta employees discussed that TikTok had more safety features for protecting minor accounts from becoming “victim discovery services.” “TikTok already

announced all U16s [children under 16 years of age] would be defaulted to private and unable to move to a more public setting” and announced that direct messaging would be defaulted to preclude direct messages for 16-17-year olds as “U16s cannot access the messaging feature.” In contrast, Meta allows minors to choose whether to make their account private: “we don’t turn off messaging by default for u16,” doesn’t offer as extensive education, “we don’t offer family pairing for teens (both youtube and tiktok are starting to have guardians supervise/monitor teens in various ways.” and TikTok “doesn’t use location . . . for u18.” Upon information and belief, Meta has still failed to make important changes, allowing, for example, direct messaging on Facebook by strangers who any other “friend” of a child user has agreed to accept as a friend and

[REDACTED]

[REDACTED]

265. In 2021, another Meta employee conveyed her concern following up on reporting by the Journal re. “preteen” hashtags,” noting that she was “continuing to find stuff that’s super concerning, not for publication at the moment, but it makes me really question whether IG is ready to build an experience for pre-teens.” She explained: “it’s comments on minors’ posts that aren’t being caught” and “will say that minor sexualization is an issue I think we need to proactively do more work in . . .” Yet, “minor sexualization” continues to plague Instagram and Meta’s other platforms.

266. As if the collective documents detailed above were not enough, in September 2021, Meta consultants, including Arturo Bejar, commissioned the BEEF study to evaluate user experiences on its platforms, to “develop a holistic, consistent picture of user bad experiences on Instagram” and create a metric that allows Meta to “track our progress” in eliminating those experiences. Meta had employed a “prevalence” metric as its preferred measurement that

ostensibly showed low percentages of offensive content appearing on Meta’s platforms. However, “prevalence” measured only that “percentage of content viewed worldwide that explicitly violates a Meta rule.” Thus, the measurement undercounted negative content because (1) it relied only on that material that Meta identified and that violated its narrowly-drawn rules; and (2) it relied on user reports as a basis for the measurement, thereby assuming that unreported illicit content (which Meta knew vastly outnumbered reported illicit content) did not violate Meta’s standards and was not otherwise improper or objectionable.⁸⁸

267. The study collected data from June 27, 2001 through July 8, 2021 and found that “[y]ounger people report higher rates of every issue, with some issues felt more universally than others.” More than half of all respondents had experienced 1 of the 22 surveyed “bad experiences” in the past 7 days, but that number is skewed by older populations having less exposure to those experiences. In fact, 54.1% of users aged 13-15 and 57.3% of users aged 16-17 had suffered a “bad experience” on Instagram in the last 7 days. A copy of the study results is attached hereto as Exhibit 1.

268. Moreover, users aged 13-15 had the highest rates of “bad experiences” for 8 of the 22 issues surveyed. 19.2% of survey respondents aged 13-15 stated they had a bad experience with “nudity” on Instagram in the last 7 days. Also consistent with the State’s investigation, 13% of survey respondents aged 13-15 stated they experienced “unwanted advances” on Instagram in the past 7 days; without the one-week limitation, that figure increases to 27% of 13-15-year olds). Results of the survey highlighting findings among 13-15 year old respondents are displayed below:

⁸⁸ <https://www.wsj.com/tech/Instagram-facebook-teens-harassment-safety-5d991be1>.

Issues by most and least impacted age groups

The 13-15 age group has the highest rates for 8/22 issues, while the 45+ group has the lowest rates of 21/22 issues

The table to the right shows the percentages of respondents who selected "Yes, during the last 7 days" for each of the issues, split by self-reported age. The green cells represent the lowest incidence for that issue across all age groups, and the red cells represent the highest incidence. Numbers that share a numeric subscript are not statistically significantly different from each other at $p < 0.05$. The full breakdown by all age groups is on the next slide.

The 13-15 age group has the most red cells (8)—that is, they have the highest rates of issue experience in 8 of the 22 issues, more than any other age group. By contrast, the oldest age group (45+) has the most green cells (21)—they have the lowest rates of issue experience in 21 of the 22 cells.

	Overall	13-15	45+
commerciality	48.2%	37.7% ^a	39.6% ^{a,f}
misinfo	30.3%	24.4% ^a	16.1% ^e
bully witness	28.3%	27.2% ^a	10.6% ^e
audience limitation	26.8%	28.2% ^{a,b}	15.9% ^e
usability passive	25.5%	30.1% ^a	10.1% ^e
hate witness	25.3%	26.0% ^a	9.0% ^e
data privacy	24.4%	17.9% ^a	14.3% ^e
perceived control	23.9%	18.4% ^a	14.2% ^d
fake acct. 1st	21.8%	15.8% ^a	13.2% ^e
negative comparison	19.2%	21.4% ^a	6.8% ^e
transparency	17.5%	14.9% ^a	8.4% ^e
political posts	17.0%	14.3% ^{a,d}	8.7% ^e
rudity	16.3%	19.2% ^a	6.2% ^e
over enforcement	14.8%	14.1% ^{a,c}	5.9% ^e
violence	12.8%	12.8% ^{a,b}	5.0% ^d
unwanted advances	11.9%	13.0% ^a	6.9% ^d
usability action	10.7%	11.2% ^{a,e}	6.6% ^d
bully target	8.1%	10.8% ^a	3.4% ^e
self harm	6.7%	8.4% ^a	1.9% ^d
acct security	3.9%	5.2% ^a	1.3% ^d
drugs	3.9%	3.7% ^{a,b}	1.7% ^c
impersonation 1st	3.7%	5.5% ^a	0.3% ^b

An expanded table with confidence intervals can be found [here](#).

20

269. These results stood in stark contrast to the rosy picture that Meta's publicly reported "prevalence" metrics painted. As the Wall Street Journal reported, "[u]sers were 100 times more likely to tell Instagram they'd witnessed bullying in the last week than Meta's bullying prevalence statistics indicated they should."⁸⁹

270. In response to the survey, Meta deployed an experiment that alerted users if their comment or post had upset other users who viewed that content. The experiment demonstrated that many users would delete the post or comment following that notification. However, Meta did not adopt that experiment or other proposed remediations. Rather, Meta personnel, including managers and even Meta's leadership (as discussed below) refused to implement changes as a result of the survey.⁹⁰

⁸⁹ <https://www.wsj.com/tech/Instagram-facebook-teens-harassment-safety-5d991be1> (last visited Dec. 1, 2023).

⁹⁰ <https://www.wsj.com/tech/Instagram-facebook-teens-harassment-safety-5d991be1> (last visited Dec. 1, 2023).

271. Frustrated by the response of company officials, Bejar sent an email directly to Meta leadership, including Zuckerberg and Mosseri, on October 5, 2021, expressing his frustrations. But, as explained below, that email did not result in further action.⁹¹

272. Adding insult to injury, Meta refused to make the BEEF survey available to other employees. Instead, Bejar was permitted only to post a “sanitized version” of the survey that omitted all of the data and approached the problem as if it were a hypothetical.⁹² Upon information and belief, this was consistent with a general corporate directive, after Haugen’s disclosures, that employees not reflect critical opinions, discussions, or analysis in documents.⁹³

273. Bejar testified before Congress in November 2023 concerning his experiences in the 2021 timeframe. He noted that, notwithstanding the alarms that he and his team raised, nothing has changed. “It is . . . two years after my briefings, and there is still no way, so far as I or teenagers I know can determine, for a minor to flag a conversation in Instagram to indicate it contains unwanted sexual advances.”

274. In February 2023, Meta Chief Executive Officer Mark Zuckerberg announced the “year of efficiency,” an effort that included a plan to layoff more than 20,000 Meta employees. Throughout the winter and spring, Meta enacted its plans, laying off thousands of workers across its workforce. Those layoffs included, and, on information and belief, disproportionately affected, content moderation and internet safety personnel.⁹⁴

275. A June 2023 Wall Street Journal report following these layoffs detailed continued troubling problems in Meta’s reporting system. An activist reported an Instagram account purporting to sell under-age content, including the phrase “This teen is ready for you pervs.” Meta

⁹¹ *Id.*

⁹² *Id.*

⁹³ *Id.*

⁹⁴ <https://www.cnn.com/2023/05/26/tech-companies-are-laying-off-their-ethics-and-safety-teams-.html>.

responded: “Because of the high volume of reports we receive, our team hasn’t been able to review this post.” Another post “of a scantily clad young girl with a graphically sexual caption” apparently did not violate Meta’s standards. Meta responded to a post reporting the account by advising the post did not violate its “Community Guidelines.” After being contacted by the Wall Street Journal, Meta acknowledged a “software glitch was preventing a substantial portion of user reports from being processed” and that its existing content moderators were not appropriately enforcing Meta’s protocols.⁹⁵

XIII. META’S BUSINESS MODEL TARGETS YOUNG USERS SPECIFICALLY WITH FEATURES DESIGNED TO ENTICE AND ADDICT YOUTH

276. Meta’s failures extend beyond the realm of disturbing CSAM, CSEC and human trafficking content existing on their platforms. At the same time that this abhorrent content proliferated, Meta was making design choices on its Facebook and Instagram platforms that were designed to capture and addict young users, including teenagers. For years, internal documents at Meta chronicled the harm caused by these features, including Meta’s pernicious algorithms. Meta was well aware that children under the age of 13 were using their platforms by lying about their age at signup, but never instituted sufficient verification technology. Again and again, Meta rejected calls from personnel to invest in its “well-being” efforts or to modify features of its platforms that were effectively proven to cause harm. Despite Meta’s public comments to the contrary, Meta and its executives failed to take sufficient action to address these known harms because such actions would have jeopardized Meta’s profits. In sum, Meta prioritized profits over safety.

277. Children were not merely unintended casualties of adult-oriented platforms that Meta did not design to keep them safe. Teenagers were, and are, a primary target for Meta, as

⁹⁵ <https://www.wsj.com/articles/instagram-vast-pedophile-network-4ab7189> (last visited Dec. 1, 2023).

demonstrated by an internal document provided to Congress by whistleblower Frances Haugen entitled “State of Teens, June 2018.” That document reports that Meta did not have “the number-one product for all use cases in all markets . . . But we do have one of the top social products—with growing marketshare—almost everywhere. Things are not great yet, but there is a visible path to greatness.”⁹⁶ As early as 2016, Mark Zuckerberg sent a lengthy email to his executive team concerning “Opportunities for Teens and Sharing.” That same year, an internal Meta email to Adam Mosseri stated “[o]ur overall goal remains total teen time spent . . . with some specific efforts (Instagram) taking on tighter focused goals like U.S. teen total time spent.” Another 2018 email stated Meta’s focus succinctly: “Short summary is the ‘the young ones are the best ones.’ You want to bring people to your service young and early.”

278. Meta’s design decisions were frequently motivated by the fact that it was losing market share among teens and adolescents to its competitors. A 2016 presentation entitled “Teens” cites concerns that teenaged users (ages 13-19) are spending more time on Snapchat and YouTube than on Meta’s platforms and proposed a plan to “win over” teenage users. Meta went so far as to put a “lifetime value” on “a 13 y/o teen” of “roughly \$270 per teen . . . Spending above that should result in much more scrutiny.” Meta thus confirmed that its interests lied in acquiring new teenage users and keeping them on the platform, but that efforts to address existing teenage users were not cost-effective and therefore discouraged.

279. In September 2019, Instagram tested an effort to make its “messaging tab” more accessible to users in order to “slow or reverse this trend” of declining teen use. That same document noted that “Teen and young adults on Instagram are responsible for 30% of US family

⁹⁶ https://www.documentcloud.org/documents/23571647-tier2_teen_ir_0618 (last visited Dec. 1, 2023).

sends and 73% of Facebook’s overall US send growth. We cannot win US messaging overall if we lose this segment.”

280. An August 2021 email stated the same concerns, noting that Instagram’s youngest users, “13 & 14 y/o” constituted “the largest components of the decline” in users, and that this trend represented “the most concerning problem from a strategic POV: they are suppose[d] to be the future of IG”

281. Having attracted teens to the platform, Meta intentionally designed its products to make sure that it maximized their time on the platforms, developing addictive-by-design features specifically targeted and tailored to exploiting, manipulating, and capitalizing young users. The net result of these features is to defeat young people’s ability to self-regulate their time spent on its platform.

282. Meta’s efforts included extensive research to understand the brain function of young people and exploit those functions to increase usage of Meta’s platforms.

283. A May 2020 presentation prepared by Meta researchers entitled “Teen Fundamentals,” discussed “adolescent development concepts, neuroscience as well as nearly 80 studies of our own product research” and highlighted vulnerabilities of the teenage brain. The presentation discussed teen brains’ relative immaturity, and teenagers’ tendency to be driven by “emotion, the intrigue of novelty and reward.” Further, and troublingly, the researchers assessed these brain functions to drive “product usage,” noting that “the teenage brain happens to be pretty easy to stimulate” and that teens’ desire for novelty “manifests itself in three behaviors that especially lend themselves to social media—exploration, discovery and experiences.”

284. In explaining “exploration,” the presentation demonstrated how Meta’s platforms attract teens’ “novelty seeking mind[s]” by “deliver[ing] [teens] a dopamine hit” every time a teen

“finds something unexpected” on the app, fulfilling their brains’ “insatiable” need for “‘feel good’ dopamine effects,” to which “teen brains are much more sensitive.”

285. The presentation noted that teens were especially prone to venturing down “rabbit holes” of viewing posts on a particular topic because of the “especially ‘plastic’” attributes of their developing brains. Further, the presentation focused on the premise that “[a]pproval and acceptance are huge rewards for teens,” and promoted the use of “[direct messages (DMs)], notifications, comments, follows, likes, etc.,” all of which “encourage teens to continue engaging and keep coming back to the app.”

286. The 2020 presentation findings were not new. A July 2018 internal presentation entitled “Problematic Facebook use: When people feel like Facebook negatively affects their life” noted “Problematic use is highest among teens and people in their 20s, consistent with previous findings that younger people generally have more problems with self-regulation.”

287. These findings back up statements reportedly made by Facebook’s founding President, Sean Parker, in 2017, who said:

The thought process that went into building these applications, Facebook being the first of them, ... was all about: 'How do we consume as much of your time and conscious attention as possible?'" "And that means that we need to sort of give you a little dopamine hit every once in a while, because someone liked or commented on a photo or a post or whatever. And that's going to get you to contribute more content, and that's going to get you ... more likes and comments." "It's a social-validation feedback loop ... exactly the kind of thing that a hacker like myself would come up with, because you're exploiting a vulnerability in human psychology. ""The inventors, creators — it's me, it's Mark [Zuckerberg], it's Kevin Systrom on Instagram, it's all of these people — understood this consciously. And we did it anyway.⁹⁷

⁹⁷ <https://www.axios.com/2017/12/15/sean-parker-unloads-on-facebook-god-only-knows-what-its-doing-to-our-childrens-brains-1513306792> (last visited Dec. 1, 2023).

288. Similarly, according to internal documents obtained by the Wall Street Journal, Meta “formed a team to study preteens, set a three-year goal to create more products for them and commissioned strategy papers about the long-term business opportunities presented by these potential users.” Preteens were and are prohibited from using Meta’s products, although (as noted below), Meta’s lax age verification procedures readily permit underage users to sign up for services by blatantly lying about their age. Yet, Meta commissioned more than a dozen studies to determine how best to tap the preteen audience and described its efforts as “big bets” on that sector of the population.

289. Meta and its employees knew that they were encouraging social media addiction among teens and adolescents. In a March 2020 chat, one employee asks for “recent studies” concerning “time spent tools.” She was directed to research Meta performed in December 2018, which established: “(1) teens feel addicted to IG and feel a pressure to be present, (2) like addicts, they feel that they are unable to stop themselves from being on IG, and (3) the tools we currently have aren’t effective at limiting their time on the app.” An internal chat from September 2020 captured two Meta employees discussing whether Meta was “creating a world of addicted monsters” in response to a documentary making that claim. Both employees agreed with the premise, stating “A lot of it is probably true” and “i would say I agree.”

290. Whether called “addiction” or another term, the outcomes and the devastating effects are the same and were the direct result of Meta’s design choices. In order to deliver this “dopamine hit” and fuel this “addiction,” Meta engineered its platforms to contain numerous features that prove irresistible to young users and addict them to the services, including: (i) engagement-based feeds; (ii) infinite scroll; (iii) push notifications; (iv) ephemeral content; (v) auto play video; and (vi) other features.

A. ENGAGEMENT-BASED FEEDS AND THE ALGORITHMS

291. When first launched, Facebook and Instagram “feeds” (the main screens where users consume content) were listed chronologically. The most recent post from an account the user followed appeared at the top, regardless of the topic or user.

292. But Meta shifted that pattern for both Facebook and Instagram. Instead of listing posts chronologically, Meta employed one or more algorithms to recommend posts and accounts to users based upon a set of criteria, including what the user had previously engaged in. The purpose of this change was to engage users for longer periods of time.

293. Meta’s algorithms are designed to interpret vast amounts of users’ data and predict what a particular user will be most interested in, thereby increasing the user’s time spent on Meta’s platforms. These predictions are presented as “recommended” content or accounts.

294. The algorithms constantly refine their predictions, measuring and analyzing user activity (including whether a user clicked on a link, or viewed a post, or otherwise interacted with the content or account) and then using those actions as data points calculated to create new recommendations. Instagram’s Adam Mosseri explained in a June 2021 blog post:

Next we take all the information we have about what was posted, the people who made those posts, and your preferences. We call these “signals”, and there are thousands of them. They include everything from what time a post was shared to whether you’re using a phone or the web to how often you like videos. The most important signals across Feed and Stories, roughly in order of importance, are:

- **Information about the post.** These are signals both about how popular a post is – think how many people have liked it – and more mundane information about the content itself, like when it was posted, how long it is if it’s a video, and what location, if any, was attached to it.
- **Information about the person who posted.** This helps us get a sense for how interesting the person might be to you, and includes signals like how many times people have interacted with that person in the past few weeks.
- **Your activity.** This helps us understand what you might be interested in and includes signals such as how many posts you’ve liked.
- **Your history of interacting with someone.** This gives us a sense of how interested you are generally in seeing posts from a particular person. An example is whether or not you comment on each other’s posts.

98

Although Meta updated this blog post on May 31, 2023, the updates did not substantively affect the information presented above.

295. The algorithm’s power lies in its ability to interpret “behavioral patterns, not by matching a user’s interests to specific subjects. This approach is efficient in increasing the relevance of recommendations.”⁹⁹ The algorithm is, in effect, a never-ending loop of future recommendations, seeking to improve the efficacy of Meta’s efforts to capture more and more of a user’s time and attention by delivering the content most likely to be engaging to them.

296. A “key psychological characteristic” of the algorithm is that it does not display content chronologically, but rather in a seemingly random pattern (random to the user at least). Thus, “even the anticipation of” the “reward” of seeing a new post or new content “can be psychologically and/or physiologically pleasing.” Indeed, “one of the main reasons why social

⁹⁸ <https://about.instagram.com/blog/announcements/shedding-more-light-on-how-instagram-works>

⁹⁹ Jeff Horwitz and Katherine Blunt, “Instagram Connects Vast Pedophile Network,” Wall Street Journal (June 7, 2023).

medial users repeatedly check their screens” is this technique, which psychologists refer to as a “variable reinforcement schedule.” “Habitual social media users never know if their next message or notification will be the one that makes them feel really good. In short, random rewards keep individuals responding for longer and has been found in other activities such as the playing of slot machines and video games.”¹⁰⁰

297. The algorithm effectively keeps “user[s] . . . in a loop” and thereby “exploit[s]” the mind’s “innate systems.” “The way this comes about is through a term referred to as Variable Reward Schedules. This works by positive stimuli being provided at random intervals” and prompts users to “check[] their phones for notifications and updates at periodic intervals for something that could be intrinsically rewarding.”¹⁰¹

298. Meta did not disclose that its algorithms were designed to leverage young users’ dopamine responses and create an addictive cycle of engagement. Nor did it disclose that the algorithm collects data in order to fuel young users’ compulsive use of Meta’s platforms, by training its algorithms to induce them to keep using Meta’s platforms. Rather, Meta makes benign claims in documents such as its Instagram Privacy Policy that it uses the data it collects “[t]o promote safety, security and integrity,” “[t]o research and innovate for social good,” or “[t]o provide measurement, analytics and business services” for individuals who “rely on our Products to run or promote their businesses.”¹⁰² These claims are intended to assuage users’ fears that their data will be used for nefarious purposes and attract or retain them as users of the Platforms, rather than to disclose the full extent of the purposes for which Meta collects data.

¹⁰⁰ Mark D. Griffiths, “Adolescent social networking: How do social media operators facilitate habitual use?”, <https://sheu.org.uk/sheux/EH/eh363mdg.pdf> (last visited Dec. 1, 2023).

¹⁰¹ Rasan Burhan & Jalal Moradzadeh, “Neurotransmitter Dopamine (DA) and its Role in the Development of Social Media Addiction,” <https://www.iomcworld.org/open-access/neurotransmitter-dopamine-da-and-its-role-in-the-development-of-social-media-addiction-59222.html> (last visited Dec. 1, 2023).

¹⁰² Privacy Policy, <https://privacycenter.Instagram.com/policy> (last visited Dec. 1, 2023).

299. However, the algorithms suffer from a fatal flaw – they do not distinguish between permissible and impermissible content. Effectively, the algorithm does not know “right from wrong.” Thus, if a young user clicks on a certain type of content—whether accidentally, out of curiosity, or intentionally—the algorithm will recommend more of that type of content.

300. This phenomenon is commonly referred to within Meta as a “rabbit hole.” Meta was acutely aware that its algorithms reinforced negative content and recommended such content to young users. These aspects were described as “deficits” in Meta’s system: “(1) it’s still very easy to find borderline and sometimes violating content/accounts in search, e.g. try typing ‘drug’ or ‘love’ [and] (2) if you start following borderline accounts or interacting with such content, our recommendations algorithms will start pushing you down a rabbit hole of more egregious content.”

301. Meta employees emailed regularly concerning “rabbit holes on potentially problematic content.” One employee noted that Meta failed to confront the problem head on. Instead of addressing the algorithms, “the primary way we combat rabbit holing you into bad stuff is by trying to detect and filter out bad stuff” but acknowledged that “we should actually be better at detecting clusters of recommendations than individual ones.” As made clear above, Meta is not effective at detecting and filtering out the “bad stuff,” leaving the algorithm to multiply its reach.

302. Meta’s algorithms are so strong that even accidental clicks triggered recommendations for harmful content. A July 2020 email discussion notes that “[w]e’ve seen over and over that many recs are generated from collaborative filter sources so if a person accidentally engaged (or baited to engage) with a non-recommendable content/entity they’ll be sucked into a rabbit hole. IG has gotten some flak for this in the press too.”

303. Indeed, even promises made to regulators concerning Meta’s algorithms were not followed through. An email chat from August 2020 recounts how, in the wake of the suicide of

14-year-old Molly Russell (an event that prompted a belated period of introspection at Meta, described below), Meta “promised the UK we’d do a bunch of work to improve IG’s impact,” including by addressing the “vortex” or “flywheel” created by the Instagram algorithms. However, more than a year after the publicity surrounding the suicide, internal emails confirmed that Meta’s “SSI [suicide and self-injury, Meta’s term for content depicting or discussing self-harm or suicide] team in Community Integrity is (slowly) working on” those promised changes.

304. Nevertheless, Meta consistently misrepresented this aspect of its algorithms as well. Antigone Davis told a congressional hearing in September 2021 that Meta unequivocally “do[es] not direct people towards content that promotes eating disorders.” The June 8, 2021 blog post entitled “Shedding More Light on How Instagram Works” represented that Instagram employed “algorithms, classifiers, and processes” in order “to personalize your experience” to prioritize “what you care about most.”¹⁰³ But that post did not disclose that those recommendations could, and often did, include recurring posts containing illegal or harmful content, like that uncovered by the State’s investigation. Nor did it disclose the existence of “rabbit holes,” “vortices” or “flywheels,” well-known to Meta, that resulted from the algorithms’ design and operation. Further, an article in Instagram’s “Help Center” section represents that Instagram “avoid[s] making recommendations that may be inappropriate for younger viewers.”¹⁰⁴ But this statement, too, ignores reality as the Attorney General’s investigation and reams of internal documents and public reports prove otherwise.

B. INFINITE SCROLL

305. Infinite scroll refers to the practice of displaying an endless stream of posts and advertisements to a user’s main feed to other sections of the platform. When a user opens their

¹⁰³ <https://about.instagram.com/blog/announcements/shedding-more-light-on-how-instagram-works>

¹⁰⁴ <https://help.instagram.com/313829416281232>

feed, they are shown a post (usually selected by the algorithm) and then a portion of the next piece of content, which the user is thereby enticed to view. This cycle continues “infinitely,” and with each post viewed, another post appears below.

306. Meta designed the “infinite scroll” feature to endlessly “tease” and/or offer new posts and advertisements for the user to view as the user scrolls down their page feed, removing any need to hit a “next page” button to view more. As a user scrolls down their feed of posts, the platform continuously and perpetually selects and shows more posts to the user.

307. The “infinite scroll” format intentionally makes it difficult for young users to leave the platform because there is no natural end point for the display of new posts and a young person is reliant solely on their undeveloped ability to self-regulate. As identified in Meta’s “Teen Fundamentals” research, this design exploits the “especially plastic” nature of teen brains to lead them down “rabbit holes.”

308. The platform does not tell a user when they have seen all the new posts from accounts they follow. Instead, the platform continues to seamlessly display and suggest additional posts from other accounts the user does not follow, provoking the young users’ well-known social “fear of missing out” on something new or interesting (commonly called “FOMO”).

309. This perpetual stream is designed to “keep [users] scrolling, and purposely eliminate any reason for [them] to pause, reconsider or leave.” The user’s experience is turned into “a bottomless flow that keeps going” and this “flow state” “fully immerse[s]” users, distorts their perception of time, and “has been shown to be associated with problematic use” of social media platforms.²⁷

C. PUSH NOTIFICATIONS

310. Facebook and Instagram also leverage “notifications” in order to entice users and increase their time spent on the platforms. Internal documents confirm that Meta viewed notifications as an essential “lever for teen growth,” as one presentation noted a key goal was to “[l]everage teens’ higher tolerance for notifications to push retention and engagement.”

311. A “push notification” is an alert displayed on a user’s device to signal that some activity has occurred on the platform and entice users—especially young users—to return to the platform and view the activity. Push notifications may be sent for a variety of activity, including when another user follows them, or likes or comments on their post. Push notifications may also appear if the user is “tagged” or mentioned in a post or if a message is sent.

312. “Sounds and vibrations are deliberately designed and distracting technologies that facilitate users’ attention away from the offline world and back to life online – pulling individuals ‘out of the moment.’” Moreover, the repeated nature of these notifications “creates a trigger for a routine and is exactly what social media operators want you to do.”¹⁰⁵

313. Meta enables “push notifications” by default when one of its apps is installed on a smartphone, and notifications may appear on a user’s screen when the phone is not being used (such as when a young user is doing homework for school) or when the user does not have a Meta app open. Notifications are not just visual; they will cause the device to vibrate and make a sound by default unless the user changes the setting. These notifications are calibrated to maximize the likelihood that a user who is not presently using the product will re-open the platform. Indeed, a small number at the top of a Meta platform icon on a user’s mobile device will display just how many notifications the user has “missed.”

¹⁰⁵ Mark D. Griffiths, “Adolescent social networking: How do social media operators facilitate habitual use?”, <https://sheu.org.uk/sheux/EH/eh363mdg.pdf> (last visited Dec. 1, 2023).

314. Meta intentionally structures the content of its notifications in order to lure users to the site. As stated in a December 2020 internal email, “[w]e limit the amount of information in notifications because we want people to come onto the site.” The email notes that notifications replaced more fulsome “emails telling you what happened,” which the employee reported was “a clear value-engagement tradeoff.” And Meta implemented its carefully designed notification system notwithstanding internal Meta documents published years earlier suggesting that Meta permit users more flexibility in choosing what notifications they receive and when because it “could help ‘twitchy’ users who may keep entering the app because of the push notifications they receive.” Another internal document suggested that Meta viewed notifications as a lever to increase growth and engagement on the platform. Meta’s “research priorities” in June 2020 regarding youth on Instagram included studying the question: “[h]ow can notifications re-engage less active users with Instagram?”

315. A recent study performed by Common Sense Media and the C.S. Mott Children’s Hospital confirms the ubiquity and intensity of notifications in a young person’s life. The research found that young users “received a median of 237 notifications” in a “typical day,” and that “[n]otification frequency varied widely, with maximums of over 4,500 delivered and over 1,200 seen.” Nearly a quarter of those notifications arrived during school hours.¹⁰⁶

316. Meta was acutely aware that sending the correct notifications in the correct volumes was a key to attracting and retaining young users. A November 2019 presentation entitled “IG Notification Systems Roadshow” confirmed: “We know notifications has an harmful effect on **teen** usage. We’ll spend H1 investigating further . . .” (emphasis in original). Nevertheless, Meta defaults young users (and all users) into receiving notifications.

¹⁰⁶ https://www.common sense media.org/sites/default/files/research/report/2023-cs-smartphone-research-report_final-for-web.pdf

317. Moreover, Meta was making its decisions with knowledge that notifications were harmful to young people. The July 2018 “Problematic Facebook use” report stated the situation bluntly: “In academic experiments, smartphone notifications caused inattention and hyperactivity among teens, and they reduced productivity and well-being.”

D. EPHEMERAL CONTENT

318. Ephemeral content is a form of content that is not permanent and disappears after a set amount of time, typically 24 hours. As one article notes, “[t]he most exciting thing about this ephemerality is the disappearance of posts.” Thus, the content “strikes a FOMO (fear of missing out) effect that creates an exclusivity to watching posts.” Such content “elicits a prompt response” and “increases user engagement.”¹⁰⁷

319. FOMO is a “pervasive apprehension that others might be having rewarding experiences from which one is absent” and “is characterized by the desire to stay continually connected with what others are doing.” Social media, including Meta’s platforms, “may be especially attractive” for individuals “who fear missing out.”¹⁰⁸ “Adolescents are particularly susceptible to development of Fear of Missing Out by using social media.”¹⁰⁹

320. Meta employs ephemeral content in order to drive user engagement by making certain content available to users only temporarily. For example, on August 2, 2016, Meta introduced a feature called “Stories” to Instagram that showed images and narratives for only 24 hours before disappearing from a user’s feed. Meta later added similar functionality to Facebook. An internal Meta document from 2018 states: “we’ve invested in FB stories—and have seen engagement more than double[:;] teen original sharing [is] up for the first time since 2012.”

¹⁰⁷ Madiha Jamal, “Ephemeral Content – The Future of Social Media Marketing,” Medium (Mar. 2, 2021), available at: <https://bettermarketing.pub/ephemeral-content-the-future-of-social-media-marketing-996d265916c2>.

¹⁰⁸ <https://www.sciencedirect.com/science/article/abs/pii/S0747563213000800?via%3Dihub>

¹⁰⁹ <https://journals.sagepub.com/doi/full/10.1177/2056305120965517#bibr1-2056305120965517>

321. Additionally, users may go “live” on Instagram and broadcast content as it happens. When a user goes “live,” the platform sends a notification, thereby advising connected users of the limited availability of the broadcast.

322. A February 2016 business plan sent to Meta’s executives, including Sheryl Sandberg, touted that Instagram’s “Live” feature was specifically intended to appeal to teens in order to maximize young users’ time on its platforms, setting goals “[t]o drive substantial watch time via Live” by “[finding] partners to appeal to teens.” There can be no question that Meta implemented “Live” in order to increase engagement among young users.

323. Because these features are only displayed for a limited time before disappearing, young users are motivated to frequently open, return to, and remain on the Instagram platform so they will not “miss out” on viewing the content before it disappears.

324. Meta employs these features even though they are aware the features cause harm to young users and that young users are unable to self-regulate their use of Meta’s platforms as the result of Meta’s design choices. An October 2019 presentation entitled “Teen Life Moments: Mental Health Deep Dive” reported that “FOMO and feeling left out” was a category of harm on Instagram, and “[y]oung people are acutely aware that Instagram can be bad for their mental health, yet are compelled to spend time on the app for fear of missing out on cultural and social trends.” Another internal document entitled “Understanding the Negative Social Comparison Spiral” noted that “FOMO” was part and parcel of activity “associated with feelings of indulgence or addiction, a behavior they get ‘sucked into.’”

E. VIDEO CONTENT

325. Displaying and “autoplaying” video content is a key aspect that Meta employs to attract and keep younger users. Videos are played in the form of “Reels,” which are short videos created by other users that are displayed in full-screen format and presented in an algorithmically-

driven feed. Like other content, Reels display “likes,” follows, comments and views of a particular video. Video represents an important part of Meta’s business strategy and of its strategy to attract youth. According to Mosseri, “more and more” of Instagram’s content will be video over time, as opposed to static pictures.¹¹⁰

326. Reels plays a key role in Meta’s business scheme to attract and addict young users. An undated presentation entitled “Teens & Young Adults on IG & FB” confirmed that Meta was “investing heavily in Reels, Stories & Creators in an effort to generate more value for teens” on a page that addressed whether teens were “joining” or “engaging” with Instagram.

327. Video clips on Facebook Reels and Instagram Reels automatically play as a user scrolls and automatically restart once they conclude, providing a stream of content without a break when young users might disengage. The videos are intentionally of a short length in order to ensure that users do not become bored. As noted above, the Reels that play are selected by Meta’s algorithm.

328. Meta launched Reels in order to attract teens who were transitioning to competitors, like TikTok, that already featured a video service. Internal Meta documents confirm that the launch of Reels was rushed in order to preserve engagement among Meta’s teen users. One employee noted in a 2020 message: “The fact that we’re shipping reels without a clear picture of the ecosystem impact is pretty mind boggling.” Another employee echoed that sentiment: “it is scary the speed we are moving . . . we either do things WAY TOO FAST without Data. Or do things WAY TO[O] SLOW because of Design/Principles.” These product designers were aware of the harm that could result from Reels, with one stating “I am worried that the cumulative effects are going to be bad.”

¹¹⁰ <https://www.forbes.com/sites/marisadellatto/2022/07/26/Instagram-exec-defends-shift-to-video-despite-complaints-from-creators-like-kylie-jenner/?sh=73be14515c6e>

329. Advertisements may appear in the midst of the videos, but, often, those advertisements are shown as videos themselves. These features encourage young users to continuously remain on the platform because they do not require user intervention to choose to view and watch the next story.

330. Meta further prominently displays its video features throughout the platform interface in order to maximize viewing time and addictive use.

F. ADDITIONAL DESIGN CHOICES

331. Additional features of Meta's platforms are designed to addict teens and adolescents.

332. Meta's platforms, and Instagram in particular, provide users with image "filters" that allow users to modify their pictures by altering their appearance in photos and videos. The filters allow facial structure alteration, body slimming, skin lightening, skin tanning, blemish cleaning, and other "beauty" features. By clicking "browse effects" when taking a picture, a user can access hundreds of these "filters," including ones named "perfect eyes" or "perfect face." Numerous filters include the word "skinny" in the title.

333. These filters significantly contribute to feelings of diminished self worth or anxiety among teens and adolescents. One study confirmed that "60% of girls feel upset when their real appearance doesn't match the online version of themselves" and professionals have expressed concerns regarding the effects of promoting "idealized and unrealistic" appearances online.¹¹¹ These effects are heightened by the fact that Meta's platforms do not identify "filtered" content to other users. Other people viewing the picture or story have no idea whether the picture reflects untouched reality or is actually digitally enhanced or altered.

¹¹¹ <https://www.forbes.com/sites/annahaines/2021/04/27/from-instagram-face-to-snapchat-dysmorphia-how-beauty-filters-are-changing-the-way-we-see-ourselves/?sh=36d5c69e4eff>

334. Meta has long been aware of the effects that cosmetic image “filters” have on teen and adolescent users, especially young female users, but has rejected or frustrated efforts to remove these filters because of fears that doing so would affect Meta’s bottom line. For example, as described in further detail below, in November 2019, Meta personnel proposed eliminating “cosmetic surgery” filters on Instagram that altered images to mimic the effects of cosmetic surgery. One employee reported that Mark Zuckerberg was skeptical that harm resulted from the filters. In response, an employee responded forcefully:

Hi,
 This email was just forwarded to me from FB colleagues. I have no strong feelings on sticky settings, but it’s been our strong recommendation from comms, marketing, policy, and engagement with nearly 20 outside experts and academics that we pass this policy which codifies a more refined and pressure tested version of our temporary policy (the one we made after the uproar last month) which disallows effects that mimic plastic surgery.

As you know, these Spark filters primarily live on IG right now and are overwhelmingly used by teen girls. Some of the arguments below seem to be rooted in ‘not giving people what they want/being paternalistic’ views but think it undermines the seriousness of the subject -- we’re talking about actively encouraging young girls into body dysmorphia and enabling self-view of an idealized face (and very western definition of that face by the way) that can result in serious issues. It’s also unrealistic to expect a large body of academic research on these subjects given the newness of the technology but the outside academics and experts consulted were nearly unanimous on the harm here.

335. Another piece of documentation confirmed that, although “robust causal evidence” was not available, “experts from around the world – APAC included – generally agree that these effects are cause for concern for mental health and wellbeing, especially amongst vulnerable populations,” including “females and “youth.” Notwithstanding this acknowledgement of harm, the email reports that Meta chose to keep the filters active on the Instagram platform, but not to “recommend” the filters to users. Meta did not choose to permanently ban the filters at the time, but only instituted a temporary ban.

336. Meta’s platforms lack sufficient parental controls and notifications. For example, the FTC’s COPPA Rule, 16 C.F.R. § 312.1 *et seq.*, requires parental notification and consent

before obtaining personal information from children. The Rule makes clear that this provision applies when the entity, like Meta, has actual knowledge that its services are used by individuals under the age of 13. As described above, Meta possesses this knowledge. However, instead of complying with this attribute of COPPA, Meta ignores its knowledge and does not require any parental consent from users, thus preventing parents from taking action if their underage child is using, misusing, or overusing Meta's platforms. Notably, Meta employs no age verification upon signup other than a user's manual input of their date of birth, and Meta is aware that underage users often use false dates of birth in order to gain access to their systems.

337. Additionally, Meta's platforms lack basic controls that would allow parents to receive notifications or warnings regarding their child's activity. Meta does not require children's accounts on Facebook and Instagram to be linked to a parent's account (or accounts). It does not send reports of a child's activity to parents, or otherwise alert parents if a child accesses illicit or concerning content (such as CSEC, CSAM, eating disorder or self-harm content for which Meta issues warnings). Meta lacks the ability to permit parents to limit a child's use, including to restrict the use of its platforms during certain hours; nor does it provide notifications to parents if a child is interacting with another adult. Meta also does not require parental consent or notification when a child follows a new account, even if that account has previously been "flagged" for improper conduct.

338. These choices are intentionally made by Meta in order to attract and retain teens. As one internal document stated, "If Mom starts using an app all the time, the app can lose a 'cool' factor" and thus Meta needed "to create spaces within the app where teens feel like they have privacy both from their own parents but also privacy from non-peers (e.g. Aunt Sally, neighbor

down the street, teachers, etc.).” That same document championed making “certain features . . . more complex (i.e. indirectly made for teens because more challenging for parents or preteens).”

339. Along those same lines, Meta made it difficult for users to self-restrict time spent on the platform. Meta offered a “time spent” tool on both Facebook and Instagram in 2018 that was described internally as “the flagship launches from IG Well-being at the time.” But the tools were faulty and ultimately discontinued. The measurements were unreliable and inconsistent across both platforms. As one Meta employee stated: “Our data as currently shown is incorrect. It's not just that Apple/Google have better data. Ours is wrong. Far worse. We're sharing bad metrics externally . . . We've been unable to right it despite several person-months of efforts.”

340. And Meta makes it difficult for users to delete their accounts if they wish to discontinue use. A user seeking to cancel their account must navigate numerous screens and pop-up messages, all of which are calculated to convince the user to second-guess their decision. Internal documents confirm that Meta's anti-deletion messages are “aggressive” and include leveraging peer pressure to convince impressionable teens and adolescents to stay, such as “list[ing] some of your friends to remind you that they will no longer be able to contact you through the site” and requiring the user to provide a reason why he or she is deactivating the account.

XIV. META WAS ACUTELY AWARE OF THE HARM TO YOUTH WELL-BEING RESULTING FROM ITS DESIGN CHOICES, BUT FAILED TO DEVOTE SUFFICIENT RESOURCES TO ADEQUATELY ADDRESS THE HARM TO YOUTH

341. At the same time that Meta was making these design choices, internal documents confirm that Meta was aware of the harmful effects that its products were having on the wellbeing of children and teenagers. Meta performed numerous studies and analyses concerning teen usage and the effects resulting therefrom, but systematically ignored internal red flags in favor of chasing profits.

342. Documents produced to Congress by whistleblower Frances Haugen confirm the tenor and pervasiveness of this internal conversation. Haugen disclosed the contents of an online, internal discussion among Meta employees in November 2016 that included messages echoing themes heard years later:

- “our ranking algorithms prevent us from truly being a blind medium . . . we are far more likely to show people only content they will agree with (in pursuit of metrics), creating an echo chamber. We are incredibly biased in what we present to people – not partisan bias, but biased towards metrics . . .”
- “[W]e only cared about things like time spent, open links, etc. That’s what we optimized for. That’s what we used to define success and failure. And that’s the problem.”

343. An internally distributed post dated August 3, 2017 is entitled “Have we made people addicted to Facebook?” The article recounts literature noting “that using apps like ours in some cases leads a user from having a habit of using the app to being addicted.” The article recommended implementing “new tools to help people not obsessively check their account,” which either were never adopted or adopted years later.

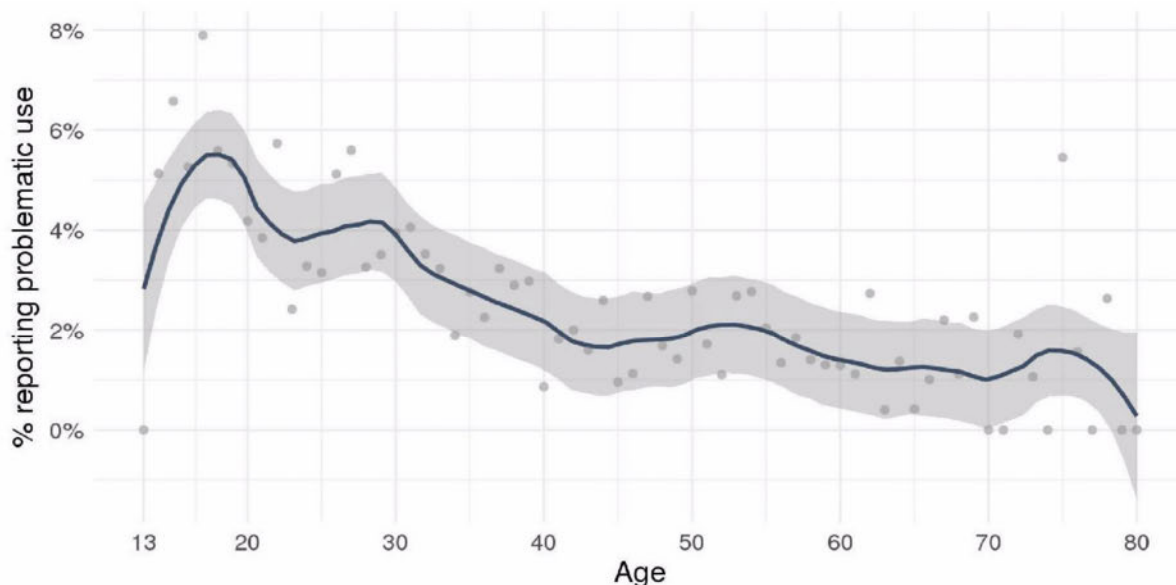
344. An undated presentation entitled “Expression & Authenticity” reported market research acknowledging the devastating social comparison aspects of Instagram: “To a great extent, social pressure is something built-in to the Instagram experience.” It notes that young people felt “[f]ears of scrutiny and judgment” when posting to their Instagram feed and “feel a pressure to be perfect on Instagram.” The presentation references “social competition” and “anxiety” among young Instagram users, as well as “fear of gossip and judgment.”

345. An internal 2018 Meta study reinforces this point, confirming that Meta has been aware for years that Instagram makes social comparison issues worse amongst teens, especially female teens:

Why? Frequent social comparison is a key driver of subjective well-being and teens say IG makes this problem worse.



346. A July 31, 2018 internal post reporting on a survey of 20,000 Facebook users is titled: “Problematic Facebook use: When people feel like Facebook negatively affects their life.” The article addresses “stories about ‘Facebook addiction,’” and identifies “activities that are most strongly associated with problematic use.” According to the findings, “teens or people in their 20s” are more likely to “feel like they have a problem” (as shown in the graph below), a finding that is “consistent with previous findings that younger people generally have more problems with self-regulation:”



Problematic use is highest among teens and people in their 20s.

347. A June 2018 internal Meta study entitled “Facebook ‘Addiction’” confirmed that, even if there was not yet evidence to support a clinical “addiction” diagnosis, Facebook could nevertheless be harmful to vulnerable users, like teens and adolescents: “Facebook may be rewarding and habit-forming,” causing “some people [to] experience genuine problems related to their Facebook use.” That presentation offered recommendations that Meta refused to implement, including limiting the number of notifications, ending the infinite scroll and reducing the availability of “Like counts:”

However, **Facebook may be rewarding and habit-forming.** Therefore some people may experience **genuine problems** related to their Facebook use.

We should reduce cases where **rewards are unpredictable or lacking in value**, and reduce **unintentional behavior**, for example:

- notifications with little or no relevance
- continuous content without breaks
- constant updates including Like counts
- lack of feedback about amount of use
- signals of availability that lead to expectations of quick responses

348. One particularly dangerous form of content that many teens sought out, had delivered to them by Meta's algorithm, and could not stop watching related to self-harm and suicide. Meta knew this but failed to take adequate steps to prevent young users from accessing this content.

349. Internal research in January 2019 underscored the presence of harmful content and the steps that teen users took to "keep one step ahead of IG's community guidelines against self-injury by changing hashtags as they are shut down," citing literature published three years earlier with titles like "Teens are using a Secret Language on Instagram to talk about Self-Harm." Indeed, under the heading "WHAT WE KNOW," the document proclaims that "Instagram is a popular platform for communicating about suicide and self-injury, and then provides shocking statistics regarding user posts in a one-month period:

SSI Desk Research

WHAT WE KNOW

1. Instagram is a popular platform for communicating about suicide and self injury.
 - a. From 7/6/18 - 8/6/18: ~16K unique IG users admitted suicidal thoughts or intent and ~21K unique users admitted self-injury. Of these, between 45% and 60% are teens, even though only 18% of DAP are teens. On a daily basis, 5.1 million users, or 0.75% of DAP, are exposed to SSI-related content each day. 5.1M users (0.75% of DAP) are exposed to SSI-related content each day (see here)
 - b. Teens keep one step ahead of IG's community guidelines against glorifying self-injury by changing hashtags as they are shut down (see Moreno et al (2016) "Secret Society 123: Understanding the Language of Self-Harm on Instagram" and Hilin (2016) "Teens are using a Secret Language on Instagram to talk about Self-Harm")

350. Tragically, despite that research, Meta continued to allow and connect young users to content that proved deadly. A 14-year-old in London named Molly Russell killed herself in November 2017 after viewing posts on Instagram and Pinterest concerning anxiety, depression,

self-harm, and suicide. Her father described the content as the “ghetto of the online world,” but, shockingly, stated in a 2022 Guardian interview that much of the content his daughter viewed before taking her own life remained accessible on Instagram even after her death and its causes became widely known.

351. The coroner’s inquest into Molly Russell’s death, published in 2022, concluded that Instagram’s algorithm contributed to her death, because the algorithm “result[ed] . . . in binge periods of images, video-clips and text, some of which were selected and provided without Molly requesting them.” The “content romanticized acts of self-harm by young people on themselves” and “sought to isolate and discourage discussion with those who may have been able to help . . . In some cases, the content was particularly graphic, tending to portray self-harm and suicide as an inevitable consequence of a condition that could not be recovered from.” Instagram was the last social media app Molly accessed before her death, and language from Instagram posts matched notes she had left in her bedroom.¹¹² The coroner concluded that the sites Ms. Russell accessed, including Instagram, “were not safe as they allowed access to adult content that should not have been available for a 14-year-old child to see.”

352. The coroner concluded that Molly died from “an act of self-harm whilst suffering from depression and the negative effects of online content.” Among the product design features that were noted during the inquest, in addition to the role of Instagram’s algorithm in providing Molly with a stream of SSI content: (1) no separation between the adult and child portions of the platforms, and content was not controlled to be age-appropriate; (2) no age verification upon opening an account; and (3) parents and guardians did not have access to the material being viewed or any control of the materials being seen by their children.

¹¹² Guardian, 9/21/2022

353. Although the coroner's inquest took several years, Meta employees were acutely aware of the lack of safeguards built into Instagram and expressed their concerns in emails following the Guardian's outreach to Meta for comments on Ms. Russell's death. In a January 26, 2019 email thread addressing Meta's response to a forthcoming media story profiling "30 families of suicide victims accusing Instagram of killing their children," one Meta employee wrote: "We are defending the status quo when the status quo is clearly unacceptable to media, many impacted families, and when revealed in press, will be unacceptable to the wider public." Recipients of the thread included Zuckerberg, Sandberg, and Mosseri. Another Meta employee responded to echo the theme that Instagram protocols were insufficient: "our present policies and public stance on teenage self harm and suicide are so difficult to explain publicly that our current response looks convoluted and evasive . . . The fact that we have age limits which are unenforced (unenforceable?) and that there are, as I understand it, important differences in the stringency of our policies on IG vs Blue App [Facebook] makes it difficult to claim we are doing all we can." Sandberg eventually chimed in, asking whether Meta could improve its policies or whether it was a question of enforcement and confirmed "We can definitely say that we need to improve our enforcement of our policies."

354. The news report concerning Molly's suicide appeared to prompt increased research into Instagram's effects on young users and surfaced internal awareness of Meta's failures with regard to SSI content. Several weeks later, an internal email to Instagram employees confirmed that "Instagram suicide and self-injury (SSI) issues reveal we are behind where we need to be." The document outlined proposed changes to Instagram's framework but noted that "because content outside of promotion did not violate our policies, we had no detection or enforcement in place at all for any other content, and most worryingly the content was easy to discovery through

search, hashtags, IG Explore, etc.” The email further noted that Instagram had not devoted sufficient resources to SSI in recent years and questioned: “Why was SSI in this state?”

355. In early March 2019, Meta employees held a “Well Being Discussion” that included Sandberg, among other attendees. Employees prepared a discussion document that confirmed that “the **average net effect** of FB on people’s well-being is **slightly negative**” and noted that 58% of Facebook users experience problematic use and 45% social comparison:

TL;DR

The impact of Facebook Inc.'s products on people's well-being is a complex topic that we haven't fully understood yet. However, we have sufficient information to warrant action now, while we continue to deepen our understanding.

1. Despite all the positive effects that FB provides (e.g. meaningful interactions, belonging, social support), there is increasing scientific evidence (particularly in the US, see [here](#) and [here](#)) that the **average net effect** of FB on people's well-being is **slightly negative**. The science here is still early, primarily US based, and relying on self report – we have a lot more we need to learn.
2. While we don't know why FB causes this negative effect, we have deep understanding around **three** negative drivers that occur frequently on FB and impact people's well-being:
 - a. **Problematic use** (prevalence: 55% mild, 3.1% severe)
 - b. **Social comparison** (prevalence: 40% mild, 5% severe)
 - c. **Loneliness** (prevalence: 36% mild, 7% severe)

356. Meta’s October 2019 “Teen Mental Health Deep Dive” presentation reports that “[o]ne in five teens say that Instagram makes them feel worse about themselves.” A particularly telling slide is entitled “Teens blame Instagram for increases in the rates of anxiety and depression among teens.” The response was “unprompted, and cited “[s]ocial comparison” among the factors that made Instagram “worse than other platforms for mental health:”

Teens blame Instagram for increases in the rates of anxiety and depression among teens

- This reaction was unprompted and consistent across all groups
- Constant comparison on Instagram is “the reason” why there are higher levels of anxiety and depression in young people
- Social comparison and perfectionism are nothing new, but young people are dealing with this on an unprecedented scale.
- The proliferation of new and different ways to compare themselves to others, combined with constant access to means that there is no way to escape social comparison on IG.
- For both boys and girls, this was called out as being the number one reason why IG is worse than other platforms for mental health. And, young people openly attribute their increased level of anxiety and depression to Instagram.

“The reason why our generation is so messed up and has higher anxiety and depression than our parents is because we have to deal with social media. Everyone feels like they have to be perfect.”

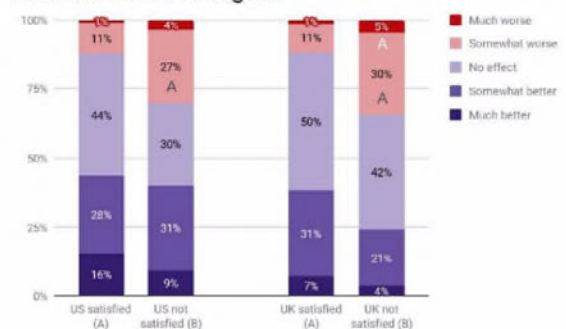
- UK Female

357. The very next slide references “fear of missing out” and includes a laundry list of “ways that Instagram harms their mental health,” including “inappropriate advertisements targeted to vulnerable groups,” and “over-sexualization of girls:”

Teens who struggle with mental health say Instagram makes it worse

- Young people are acutely aware that Instagram can be bad for their mental health, yet are compelled to spend time on the app for fear of missing out on cultural and social trends.
- Teens specifically call out the following as ways that Instagram harms their mental health:
 - pressure to conform to social stereotypes
 - pressure to match the money and body shapes of influencers
 - the need for validation -- views, likes, followers
 - friendship conflicts, bullying, and hate speech
 - over-sexualization of girls
 - inappropriate advertisements targeted to vulnerable groups

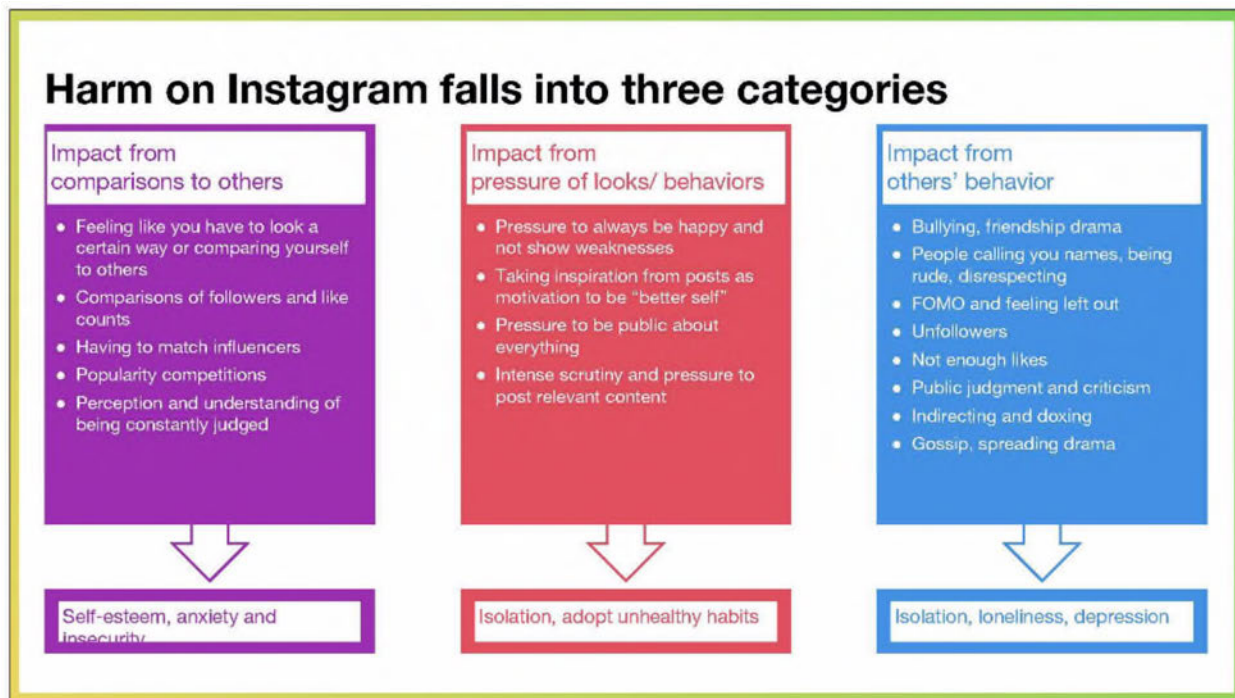
Stated effect of Instagram



Q: In general, how has Instagram affected [the way you feel about yourself/your mental health]?

There were no statistically significant differences among those who answered for “the way you feel about yourself” and those who answered for “your mental health”
US n = 1298; UK n = 1308

358. Teens further relayed that harm resulting from Instagram use “falls into three categories” and leads to devastating conditions, such as “anxiety,” “insecurity,” “isolation,” loneliness,” and “depression:”



359. Yet Meta employees did not disclose these, or similar metrics, to the public. For example, Meta refused to provide statistics to the media that would disclose to the public the true extent (or lack thereof) of their wellbeing enforcement efforts. In June 2019, a reporter for Time sent an email with follow-up questions that included inquiries regarding the number of employees “working on the issue of bullying.” A Meta employee did not want to answer the question, likely because the personnel count was paltry. The employee confirmed that “[t]he Well-being Bullying team is ~15 people [at Instagram]. The Facebook Bullying and Harassment team is about that again.” The entire Well-being team included 150 employees, but they were responsible for “all the account and content problems and elections etc, well beyond Bullying.” Meta responded to Time by identifying “30,000” individuals, a number that was comprised almost entirely of outsourced

content reviewers who did not have specific responsibility for bullying issues. Meta’s Annual Report filed with the SEC for the year ending December 31, 2019 claimed that “[a]s of December 31, 2019, we had 44, 942 employees.”¹¹³

360. An internal post disclosed by Frances Haugen dated November 5, 2019 entitled “Hard Life Moments – Mental health deep dive” confirmed that “~30% of teen girls . . . felt that Instagram made dissatisfaction with their body worse.” Additional slides contained titles such as “Social Comparison has High Reach, Moderate Intensity and We are Making It Worse For 1 In 4 People” and “Body Image (A Related Issue) Also Stands Out For Teens.” The latter slide reports that “1 in 3 teen girls blame Instagram for making their body image issues and problematic social media use worse. Social Comparison is a high reach, high intensity issue that 1 in 5 teens thought we made worse.”

361. That same message was distributed to Meta employees—but not publicly—in a February 2020 “Community RYSK (Research You Should Know)” post: “Teenage girls are 8x more likely to engage with negative social comparison (NSC) than adult men on IG. 1 in 3 teen girls report IG makes their body image issues worse. NSC can cause or is correlated with other issues around: body image, self-esteem, loneliness, and depression.” An internal Meta “literature review” dated March 13, 2020 similarly reported that “33% of Instagram users and 11% of Facebook users think the platform makes their own body image issues worse.”

362. Internal posts and research acknowledged (but not publicly) bullying on Instagram, as well. An undated presentation that likely was prepared before or in 2020 states “We knew that people experience unwanted interactions involving random strangers and people they are relatively unconnected with” and that “7% of teens report experiencing bullying on Instagram *each week*.”

¹¹³ https://s21.q4cdn.com/399680738/files/doc_financials/2019/ar/2019-Annual-Report.pdf at p. 9.

(emphasis added). The presentation noted that “B&H [bullying and harassment] prevalence in DM’s is very high (10%)” and “Severe violations targeted to those under 18 are known as a high prevalence problem in DMs.” Yet, Meta CSERs reported a prevalence number for bullying and harassment in only 8 of 10,000 views, or .08%.

363. Internal documents make clear that, despite all of these internal studies demonstrating harm from its platforms, Meta refused to devote sufficient resources in order to address the problems, despite its public statements to the contrary. An August 2020 email cited a “severe lack of capacity for restricted content,” including “nudity, graphic violence, child safety and SSI content review:”

Lowlights

- **Multiple problem areas** continue to have a severe lack of capacity for restricted content. We’ve continued to struggle with getting capacity for nudity, graphic violence, child safety and SSI content review due to restrictions on reviewing these categories of content from home. We’ve recently kicked off a CIPGO taskforce to better correlate demand and supply on restricted content, and better articulate impact on our metrics and CSERv7.

364. Meta employees internally questioned lax enforcement with respect to policing content related to eating disorders. In September 2020, a Meta employee reported disturbing deficiencies in detection and removal of information related to eating disorders. The employee reported that Meta blocked content when a user searched using known “inherently violating hashtags,” but that the search still returned results for accounts—the same problems noted with respect to CSAM and other sexually explicit content. “[W]hen you search for these terms, there are no results under the ‘hashtag’ tab, but there are endless results under the ‘Top Accounts’ and ‘Accounts’ tab, and almost all are violating.” Moreover, the same employee reported “No ED proactive detection at the moment: We don’t currently have classifiers to catch eating disorder proactively – which means it’s easy to find on Instagram, and is showing up in Explore and accounts ‘suggested for you.’” Nevertheless, a Product Manager in Instagram Wellbeing

confirmed the volume of improper and harmful content related to eating disorders: “[I]t’s hard to defend why we’d block certain terms in hashtag search and not in account search, and *it’s easy to find violating ED accounts in a couple of taps*. When I search for ‘thinspo’ [referencing “thinspiration”] or ‘proana’ [pro-anorexia] *the accounts shown are overwhelmingly violating.*” (emphasis added). Yet, Meta never disclosed these results to the public, instead relying on its misleading statements proclaiming its platforms as “safe” and claiming that it prioritized safety.

365. Moreover, on information and belief, Meta has the ability to track how much time teens and adolescents are spending on content associated with eating disorders. Meta has the technological acumen to send teens and adolescents notifications if they are spending inordinate time periods browsing this content, or to notify parents of such behavior. Meta has chosen not to take any of these steps.

366. Further discussion in the same September 2020 email thread suggested that “ED content is too narrowly defined by our policies so we under-enforce on what the broader public might consider to be ED content” -- the same deficiencies observed by the State with respect to CSEC. Emails from November 2021—more than a year later—confirm the issue had not been resolved. A journalist reported an account of a man “contacting teenagers affected by eating disorder.” The account “has only shared 2 posts but they both show extremely thin bodies, one of which has a caption #thighgap #skinny #weightloss” and “[t]he profile picture is also zoomed image of a rib cage and the ‘bio’ says ‘coaching.’” Moreover, “many of the accounts he is following and that are following him are ED-related accounts.” Despite having access to the account and its posts, Meta’s “safety team” concluded the account “had no violations.” Following that determination, the reporter explained to Meta that the “account is one of many ‘coaches’ – often with paedophile interests or a skinny fetish – who contact girls on IG and other messaging

apps and offer to ‘coach’ them in their anorexia and extreme weight loss, often soliciting photos from them at the same time.” Only after receiving that “additional context,” the account was “re-review[ed]” and Meta deleted only one of two posts for violating its Community Standards. The email shows no indication that the account was suspended or that Meta took any other action concerning these troubling reports.

367. An undated document entitled “Eating Disorders PRI Analysis Document” further documented issues regarding eating disorder content on Instagram. As with the Attorney General’s investigation of CSEC, the author created a “test account” who followed content related to eating disorders and found “[m]y test user was served violating content and accounts through IG explore surface and Accounts Suggested For You after following other ed [eating disorder]-related accounts.” The author reported that “Violating accounts and content for eating disorders were recommended through Suggestions For You and IG Explore” and explained the reasons for these recommendations included: “We rely on ED strikes from user reports within a 90 day timeframe to be classified as not recommendable” and violative “content is not being deleted” because “[w]e don’t have proactive detection for eating disorder violations in all markets.” The document further notes that “Auto Detection doesn’t exist” for eating disorder content and that “[u]sers can create and maintain IG Story Highlights that are named violating terms,” which “would not be removed off the platform.”

368. Another email in April 2021 noted multiple areas of deficiency, stating “[b]eyond teen safety defaults and reporting flow for under 13, we need solid solutions for detecting, measuring, and enforcing against children/teen safety and Well-being issues across Messenger Kids, IG, and IG Youth.” The email noted “key risks/issues” that mirrored many of the topics addressed above, including “bullying & harassment” and “age verification,” and noted the “biggest

issue” remained “lack of capacity,” i.e. Meta’s refusal to invest in personnel and resources to address these problems:

Some of the key risks/issues highlighted by the team are:

1. We do not have a specialized focus on minors within Bullying & Harassment (B&H).
2. We do not really have any bullying & harassment coverage for messenger and MK .
3. To the team’s awareness, we do not have a specialized focus on minors within any other problem areas in Central Integrity.
4. Age verification (for under 13) has a big backlog and demand is outpacing supply.
5. Age verification has no ground truth to give robust measurement of effectiveness. Some reviews lack market specialization.
6. We do not have focused efforts to understand/address adults compromising the accounts of minors to facilitate abuse, including adversarial evasion of age prediction.
7. We are very nascent in developing strategy for E2EE, ePD and other potential privacy regulations.
8. Perhaps the biggest issue: progress on these and any other issues not raised is hampered by lack of capacity. Indeed, we didn’t dwell so much on gaps on existing areas of coverage but they are

substantial as well. For example, Central Integrity has 1 DS on Child Safety and we are working to get a bit more support, but addressing gaps will require a more robust XFN effort. For plans to build a new, youth-oriented app, that (like any another new surface), will need dedicated integrity focus that does not exist today.

369. Internal Meta research from March 2021 confirmed that a shocking 33% of material viewed by teen girls on Instagram is “likely” to make them “feel worse about their bodies” and that the majority of that content was attributable to influencers and “content from non-friends:”

Takeaways

- Some topics are likely to make viewers, especially teen girls, feel worse about their bodies:
 - Fashion and beauty (especially makeup, skincare, hair removal, nail art)
 - Relationships (which includes dating, romance, sex, and intimacy)
 - Western pop stars who emphasize their bodies (e.g., Selena Gomez, Cardi B, Ariana Grande)
 - Images that emphasize women's bodies generally (e.g., cleavage, swimwear)
- Collectively, these topics comprise 1/4 of what people see on Instagram (1/3 for teen girls).
- On the other hand, seeing content that emphasizes a collective identity (e.g., sports fandom, tragedy, religion, politics) or an activity (crafts, gardening, puzzles, TV) rather than the poster's body, is associated with less appearance comparison.
- Appearance comparison is being driven by top accounts and content from non-friends. When friends post about problematic topics (e.g., beauty), it doesn't trigger appearance comparison the same way content from non-friends and celebrities does.
- For every piece of friend content a teen girl sees, she sees 5x as many pieces of content from top accounts.

370. None of these internal emails prompted Meta to address its internal deficiencies, despite its public statements to the contrary. Meta's efforts to address youth well-being and prevent harm remained under-resourced through at least August 2021. In an internal email, one Meta employee wrote: "We are not on track to succeed for our core well-being topics (problematic use, bullying & harassment, connections, and SSI, and are at increased regulatory risk and external criticism." That employee noted that "Problematic Use on FB/IG" was "the topic where we have the largest gap relative to competitors and external expectations as well as minimal current staffing" and "recommend[ed] at a minimum staffing these up for 2022."

371. None of the reports and information cited above prompted Meta to make substantial changes in its platform to address these issues. Nor did Meta disclose any of the harm it had

identified internally to the public, which would have corrected the misleading and deceptive nature of its public statements proclaiming its platforms “safe.” Rather, Meta continued to pursue profits over safety even when its own employees were identifying the harms resulting from Meta’s own design choices. As one Meta employee put it following publication of leaked internal research, “our own research confirmed what everyone has long suspected.”

XV. DESPITE KNOWLEDGE OF THE HARMS CAUSED BY ITS PLATFORMS, META REPEATEDLY REJECTED EFFORTS TO IMPROVE ITS SYSTEMS

372. Notwithstanding the abundant internal research demonstrating harms to young users of its platforms and its public commitments to make Meta safe and good for children, Meta callously rejected numerous proposals to implement changes that would have improved users’ well-being.

373. In 2019, a discussion was held concerning “what it would take to 10x well-being, which resulted in a plan that was reviewed with Chris Cox, Meta’s Chief Product Officer, and Sheryl Sandberg and then was submitted to Mark” Zuckerberg, who refused additional funding for the request. As a result, the effort “ended up punted given other ambiguities.”

374. These decisions were referenced in an August 25, 2020 internal message in which Meta employee [REDACTED] announced her departure from Meta and voiced her frustrations regarding executive decisions. Ms. Hunzaker’s email referenced “many hard integrity battles fought,” and made clear that, despite its claims, Meta did not value safety and children’s well-being, and instead subordinated, rejected, or rolled back efforts to address those concerns for monetary or other reasons. Ms. Hunzaker’s email confirms that “Integrity teams are facing increasing barriers to building safeguards” because safety efforts were “prematurely stifled or severely constrained by key decision makers,” and that Meta was “roll[ing] back” “functioning safeguards” because of “fears of public and policy stakeholder responses:”

Some parting thoughts...

1. Integrity teams are facing increasing barriers to building safeguards.

- In recent months, time and again I've seen promising interventions from integrity product teams, with strong research and data support, be prematurely stifled or severely constrained by key decision makers—often based on fears of public and policy stakeholder responses.
- Similarly (though even more concerning), I've seen already built & functioning safeguards being rolled back for the same reasons (e.g., see [sparing sharing deprecation](#), and movement back towards [lighter misinfo UI treatments](#))
- While mountains of evidence is (rightly) required to support a new intervention, none is required to kill (or severely limit) one.
- *-*To be clear - This is in **no way** meant as a criticism to the integrity teams on the ground who are working and innovating as hard as they can under increasing constraints. This (and the following) **is** intended as a call to reflection for those decisions-makers imposing constraints.*

375. She further confirmed that Meta was “*knowingly* exposing users to risks of integrity harms” because of “potential” media responses. She confirmed that cross-platform enforcement of Meta’s rules and policies was non-existent because “there is no dedicated team or central body tasked with monitoring” Meta’s “complex system . . . with cross-surface interactions for safety risks as a whole.” And she noted that Meta lacked “consistent resources or incentives provided to help teams check for potential user harms or de-risk products at the outset.”

376. Recent congressional testimony from Arturo Bejar confirms Hunzaker’s experience. Bejar initially worked in “Facebook’s Protect & Care group” until 2015, and when he

left Facebook, he “felt good that we had built numerous systems that made using our products easier and safer.”

377. Bejar returned as a consultant to Meta in 2019 to address the “distressing experiences” faced by many teens he knew of. He discovered that his prior work from just four years earlier had been largely dismantled:

It was not a good experience. Almost all of the work that I and my colleagues had done during my earlier stint at Facebook through 2015 was gone. The tools we had built for teenagers to get support when they were getting bullied or harassed were no longer available to them. People at the company had little or no memory of the lessons we had learned earlier.

378. Meta had effectively abandoned his prior work for the sake of pursuing greater teenage audiences, which in turn led to greater profit. As Bejar noted in his written Congressional testimony, “Meta’s current approach to these issues only addresses a fraction of a percent of the harm people experience on the platform. In recent years, repeated examples of harm that has been enabled by Meta and other companies has come to light, through whistleblowing, outside research studies, and many stories of distressing experiences people have there . . . [T]here is a material gap between their narrow definition of prevalence and the actual distressing experiences that are enabled by Meta’s products.”

379. Bejar laid the blame squarely on Meta and its decision-makers, including Zuckerberg: “[M]anagers including CEO Mark Zuckerberg do not seem to seek to understand or actually address the harms being discussed. Instead, they minimize or downplay published findings, and even sometimes the results of their own research. They also try to obfuscate the situation by quoting statistics that are irrelevant to the issues at hand.”

380. As another example of its unwillingness to favor young users’ mental health over its advertising revenue, Meta rejected a program intended to hide the public display of “likes”

despite knowledge that publicizing the counts of likes on users' posts caused harmful negative social comparison in teens.

381. "Likes" are another aspect of Meta's platforms that provide "[s]ocial affirmation and validation." As one research report noted, "such a simple characteristic has reaped huge rewards in terms of adolescents repeatedly coming back to check their social media platforms, and what some have described as a 'craving for validation.'"¹¹⁴ By default, Meta reports the total number of likes every post receives.

382. Meta's internal research confirmed that this public display of likes was inextricably linked with negative social comparison among young users. An email dated October 1, 2019 reported that hiding "likes" was discussed at Meta as early as December 2017 because a public roundtable "revealed how big of [a] negative impact like counts were having, especially on young creators" and noted "a broader sense that Instagram has created a #FOMO culture and a need to be perfect," resulting in "an apparent corresponding impact on mental health/anxiety levels." Citing "past research," an October 2020 internal document entitled "Likes and Social Comparison on Instagram" confirmed that "Like and comment counts have been cited as a potential trigger of social comparison. In particular, a 'Like paradox' (i.e. that people's friends tend to receive more Likes than they do) may lead people to overestimate their friends' popularity and thus feel worse." That research paper confirmed: "Seeing more posts with high Like counts (1k+) was associated with feeling worse: more negative social comparison and less positive social comparison."

383. To address this harmful impact, Meta implemented a test program called "Project Daisy," which would have hidden the public like count on posts. There are two versions of "Project Daisy:" (1) "Pure Daisy" (like counts on all posts except one's own were hidden), and (2) "Popular

¹¹⁴ Mark D. Griffiths, "Adolescent social networking: How do social media operators facilitate habitual use?", <https://sheu.org.uk/sheux/EH/eh363mdg.pdf>

Daisy” (like counts on posts from certain highly followed accounts were visible, but the like counts on average users’ posts were hidden). Both projects were well-received among young users, and Meta’s researchers recommended implementing Project Daisy in October 2020.

384. Even though “Pure Daisy reduced negative social comparison in teen girls, . . . the impact of negative social comparison was not evaluated in the original launch decision.” Instead, Meta employees, including, on information and belief, top leadership (Zuckerberg and Mosseri), rejected rollout of “Project Daisy” because “it was not impactful from the lens that IG was evaluating previously (reduce creation flow abandonment and reducing deletion of posts with low like counts) and was pretty negative to FB metrics.”

385. However, in public statements, Meta misrepresented its reasons for refusing to implement Project Daisy. Its prepared talking points regarding Project Daisy stated that Meta was implementing Daisy as an opt-in feature because “[f]or some people, hiding public like counts helped people focus less on the number and more on the content, [but] for others it didn’t really matter much.” Mosseri falsely stated in a public interview that implementing the project didn’t have “a big impact on how much people use the product” and that the impact was “neutral.”

386. Despite Meta’s outward-facing public message that it prioritized safety and the well-being of its users, the rejection of “Project Daisy” demonstrates that, in practice, user safety and well-being was subordinated to business concerns, including profits. One employee called the approval process “awful, with leads disagreeing about whether social comparison survey results counted as ‘wellbeing’ and generally discounting” the Integrity team’s “expertise.” The employee noted that Project Daisy was rejected even though “there were small but probably non-trivial revenue impacts” and despite, one employee’s words, “one of the clearest things (supported by research) that we can do to positively impact social comparison and well-being on [Instagram].”

387. In October 2020, Adam Mosseri emailed Instagram’s “Wellbeing” personnel to advise them that he believed “social comparison [is] the existential question Instagram faces within the broader question of whether or not social media is good or bad for people,” but he recognized in that same email that it would “be impossible” for the team to devote sufficient resources “without divesting from somewhere else.”

388. Yet, nearly a year later, the “Wellbeing” team remained under-resourced and subordinated to other priorities. An internal chat from September 2021 underscored the features that were lacking on Instagram compared to its competitors. Pavni Diwanji, a Meta Vice President, advised Mosseri that the “state of mental wellbeing on IG needs work.” In particular, Diwanji criticized Instagram’s “one size fit all users when it comes to mental wellbeing” and Instagram’s lack of “user visible features” available on its competitors’ products, specifically noting features that Meta rejected/refused to implement, such as “take a break, screen time management, [and] restricted mode where they allow parents to pair with teens and tweens.” Diwanji further noted that TikTok was “more proactive on content takedowns.” Diwanji, who, on information and belief, was responsible for youth initiatives at the time, complained about “challenges inside” Meta that made her job harder and warned: “this is not a good way to operate.”

389. Meta proclaimed to the public that safety was a “top priority” that it “cared deeply” about and “aggressively” worked to promote. Yet its business decisions told a different story. Time after time, Meta’s executives rejected safety and refused to invest in programs or personnel sufficient to address Meta’s significant “well-being” flaws, because doing so would have compromised Meta’s revenues.

XVI. META’S PLATFORM DESIGN CAUSED MENTAL HEALTH HARM TO YOUNG USERS, INCLUDING USERS IN NEW MEXICO

390. The design of Meta’s platform and its decisions to implement (or not to implement) features in order to maximize teen engagement and profits caused real and lasting harm to young users, including young users in New Mexico.

391. In May 2023, the Surgeon General of the United States issued an advisory entitled “Social Media and Youth Mental Health” that summarized findings concerning the devastating impacts of social media use and noting “increasing concerns among researchers, parents and caregivers, young people, healthcare experts, and others about the impact of social media on youth mental health.”¹¹⁵ While noting benefits from social media use, the advisory warned “that adolescents who spent more than 3 hours per day on social media faced double the risk of experiencing poor mental health outcomes including symptoms of depression and anxiety.”¹¹⁶ The Surgeon General further explained that studies on “college-aged youth” had shown “sizable effects,” including stark increases in depression and anxiety, and “raise serious concerns about the risk of harm from social media exposure for children and adolescents who are at a more vulnerable stage of brain development.”¹¹⁷

392. The Surgeon General’s concerns were not limited merely to time spent on social media platforms. The advisory warned that “[e]xtreme, inappropriate, and harmful content continues to be easily and widely accessible by children and adolescents,” and noted studies finding that “discussing or showing this content can normalize such behaviors, including through the formation of suicide pacts and posting of self-harm models for others to show.”¹¹⁸ The Surgeon

¹¹⁵ Advisory, <https://www.hhs.gov/sites/default/files/sg-youth-mental-health-social-media-advisory.pdf> at 4 (last visited Dec. 1, 2023).

¹¹⁶ *Id.* at 6.

¹¹⁷ *Id.* at 7.

¹¹⁸ *Id.* at 8.

General further cited studies “demonstrat[ing] a significant relationship between social media use and body image concerns and eating disorders, with social comparison as a potential contributing factor.”¹¹⁹

393. The Surgeon General confirmed that “[e]xcessive and problematic use of social media can harm children and adolescents by disrupting important healthy behaviors” and warned that product features, like those implemented by Meta, “designed to maximize user engagement . . . [have] the potential to encourage excessive use and behavioral dysregulation [typically referring to anxiety, depression, suicidal thoughts, self-harm, and other self-damaging behaviors].”¹²⁰ The Surgeon General also relayed statistics indicating that “one-third or more” of the youngest users (girls aged 11 to 15) “say they feel ‘addicted’ to a social media platform.”¹²¹

394. A systematic study of research papers printed in 2020 validated much of the Surgeon General’s analysis, concluding that “[s]ocial media are responsible for aggravating mental health problems” and finding “a general association between social media use and mental health issues.”¹²² The review noted links between increased usage of social media and “anxiety and depression,” including among teens, who “experience anxiety from social media related to fear of loss, which causes teens to try to respond and check all their friends’ messages . . . on a regular basis.”¹²³

395. Another study found that “engagement with photo-based social media sites, such as Instagram, is associated with poor body image.”¹²⁴ That study assessed social media behaviors, including “avoidance of posting selfies, photo investment, photo manipulation, and investment in

¹¹⁹ *Id.*

¹²⁰ *Id.* at 9.

¹²¹ *Id.*

¹²² <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7364393/> (last visited Dec. 1, 2023).

¹²³ *Id.*

¹²⁴ <https://onlinelibrary.wiley.com/doi/epdf/10.1002/eat.23256> (last visited Dec. 1, 2023).

others' selfies" and found that each behavior "was associated with greater likelihood" of suffering from eating disorders,¹²⁵ which can also lead to other health consequences, such as gastrointestinal illnesses, impacts to the endocrine and cardiovascular systems, bone or gray matter brain loss or atrophy, and fertility issues.¹²⁶

396. Indeed, as early as 2014, a study entitled "NetTweens: The Internet and Body Image Concerns in Preteenage Girls" concluded that "[t]ime spent on . . . social networking sites produced stronger correlations with body image concern than did overall Internet exposure" and that "the Internet represents a potent sociocultural force among preteenage girls."¹²⁷ A 2018 open letter to Zuckerberg signed by 118 public health advocates cited this study and others in concluding that "a growing body of research demonstrates that excessive use of digital devices and social media is harmful to children and teens."¹²⁸

397. Research has linked excessive social media use with sleep disturbance, which is particularly a problem among teens and adolescents and can result in "a range of poor health outcomes," including adverse effects on "cognitive performance, mood, immune function, cardiovascular risk, weight, and metabolism."¹²⁹ The researchers found "consistent, substantial, and progressive associations between [social media] use and sleep disturbance" among young adults, a finding likely equally applicable to teens and adolescents.¹³⁰

¹²⁵ *Id.*

¹²⁶ *See, e.g.,* Anorexia Nervosa, Cleveland Clinic <https://my.clevelandclinic.org/health/diseases/9794-anorexia-nervosa#outlook--prognosis>; Bulimia Nervosa; Cleveland Clinic <https://my.clevelandclinic.org/health/diseases/9795-bulimia-nervosa#symptoms-and-causes> (last visited Dec. 1, 2023).

¹²⁷ <https://journals.sagepub.com/doi/epub/10.1177/0272431613501083> (last visited Dec. 1, 2023).

¹²⁸ <https://fairplayforkids.org/wp-content/uploads/archive/devel-generate/gaw/FBMessengerKids.pdf> (last visited Dec. 1, 2023).

¹²⁹ <https://www.sciencedirect.com/science/article/abs/pii/S0091743516000025> (last visited Dec. 1, 2023).

¹³⁰ *Id.*

398. Statistics from the CDC’s “Youth Risk Behavior Survey” confirm that the rise in social media use among teenagers and adolescents, including use of Instagram and Facebook, corresponds with a decline in youth mental health. According to the survey, “[i]n 2021, 42% of high school students felt so sad or hopeless almost every day for at least two weeks in a row that they stopped doing their usual activities.” The percentage of female high school students who reported feeling “persistent feelings of sadness or hopelessness” increased from 36% in 2011 to 57% in 2021. Twenty-nine percent of all respondents and 41% of female respondents reported experiencing “poor mental health” in the past 30 days. And the survey reported that 30% of female high school students had “seriously considered attempting suicide” during the past year, an increase from only 19% 10 years earlier.

399. A 2019 study of nearly 7,000 adolescents found that “adolescent social media use was prospectively associated with increased risk of” adverse mental health characteristics. The study found that “Adolescents who engage in high levels of social media use may experience poorer quality sleep” and that increased social media use could be associated with an increased risk of “cyberbullying, which has a strong association with depressive symptoms.” Further, the study noted “negative body image,” “anxiety” and “depression” as connected to social media use.¹³¹

400. The harms described in the Surgeon General’s advisory and the representative research cited above are not limited to a particular geography. They occur nationwide, including in New Mexico.

¹³¹ Kira Riehm et al., *Associations between time spent using social media and internalizing and externalizing problems among US youth*, 76(12) JAMA Psychiatry (2019), <https://jamanetwork.com/journals/jamapsychiatry/fullarticle/2749480> (last visited Dec. 1, 2023).

401. Moreover, the Surgeon General’s advisory and public research is consistent with years of internal Meta research chronicling harm to young users arising from the design of Meta’s platforms, much of which is described in the preceding paragraphs. While Meta promised to safeguard the health and safety of children on its platforms (and to keep the youngest users offline), at every turn, it made decisions that put its own profits ahead of their well-being.

XVII. META MADE MISLEADING AND DECEPTIVE CLAIMS IN PUBLIC STATEMENTS THAT ITS PLATFORM WAS SAFE OR THAT IT WAS ADDRESSING PROBLEMATIC CONTENT

402. As Bejar made clear in his Congressional testimony, “Meta continues to publicly misrepresent the level and frequency of harm that users, especially children, experience on the platform.” These misrepresentations were designed to increase usage of Meta’s platforms by assuring teens and their parents that its platforms were safe and appropriate for children, and failed to disclose the evidence of serious harms that it knew its young users suffered. Each of these statements and omissions painted a misleading and untrue picture of the safety of its platforms and sought to minimize or misstate the volume of objectionable content and dangerous activity on Meta’s platforms.

403. Meta and its executives repeatedly made statements intended to reassure users, advertisers paying to reach those users, and parents that its platforms are safe and that illegal content on Meta’s platforms is minimal, or, if it does exist, is subject to prompt and effective action by Meta’s personnel and automated detection systems, including the following:

- a. In 2012, a Facebook spokesperson told the Associated Press that Meta “take[s] human trafficking very seriously and a number of measures are in place to counter this activity.”
- b. Eva Chen, a Meta employee responsible for fashion partnerships, was reported as stating at a CornellTech event in the Spring of 2018 that “[m]aking the

community a safer place, a place where people feel good, is a huge priority for Instagram . . . I would say one of the top priorities.”¹³²

- c. In a June 2019 interview with CBS News, Instagram head Adam Mosseri stated: “We’ve been focused on well being broadly, like I said, it’s our number one priority” and “[w]e will do things that mean people use Instagram less if we think that they keep people safe or generally create a healthier environment.”¹³³
- d. On a public earnings call on January 29, 2020, Meta’s then Chief Operating Officer, Sheryl Sandberg, stated: “While we continue to invest in helping businesses, we are equally focused on keeping our platform safe.”
- e. In an April 2020 public SEC filing, Meta claimed: “We use sophisticated technology and other techniques not only to detect child exploitation imagery and remove it, but also to detect and prevent grooming or potentially inappropriate interactions between a minor and an adult.” Meta further stated: “We deploy technology across all of our platforms to proactively surface as much illegal child exploitative content as we can, including through detection technology, machine learning and artificial intelligence techniques, and open-sourcing photo- and video-matching technology.”¹³⁴

¹³² <https://qz.com/quartz/1238074/Instagrams-new-wellbeing-team-will-address-its-effect-on-mental-health/>.

¹³³ <https://www.cbsnews.com/news/adam-mosseri-instagram-is-seriously-considering-hiding-likes-apps-head-reveals/>.

¹³⁴ Meta Proxy Statement (Schedule 14A) at 79.

- f. In May 2021, Mosseri told reporters that Instagram’s effects on the wellbeing of teenagers and adolescents was “quite small,” according to a September 2021 report in the Wall Street Journal.¹³⁵
- g. In June 2021, Meta issued a statement to CBS News calculated to portray Meta’s platforms as safe and to convince uses and potential users that Meta actively and effectively polices activity on its Facebook platform:

Sex trafficking and child exploitation are abhorrent and we don’t allow them on Facebook. We have policies and technology to prevent these types of abuses and take down any content that violates our rules. We also work with safety groups, anti-trafficking organizations and other technology companies to address this and we report all apparent instances of child sexual exploitation to the National Center for Missing and Exploited Children.
- h. On October 5, 2021, Meta Chief Executive Officer Mark Zuckerberg published a public Facebook post that claimed: “We care deeply about issues like safety, well-being and mental health. It’s difficult to see coverage that misrepresents our work and our motives.” Zuckerberg continued: “At the heart of these accusations is this idea that we prioritize profit over safety and well-being. That’s just not true.” He claimed that “it’s very important to me that everything we build is safe and good for kids.”¹³⁶
- i. Instagram’s website contains a blog post dated December 7, 2021 attributed to Mosseri entitled “Raising the Standard for Protecting Teens and Supporting Parents Online.” That post begins with the following statement: “At Instagram,

¹³⁵ <https://www.wsj.com/articles/facebook-knows-Instagram-is-toxic-for-teen-girls-company-documents-show-11631620739>.

¹³⁶ <https://www.facebook.com/zuck/posts/10113961365418581>.

we’ve been working for a long time to keep young people safe on the app.”¹³⁷

Another blog post from December 15, 2022 is entitled “Continuing to Keep Instagram Safe and Secure,”¹³⁸ which is intended to deemphasize the prevalence of inappropriate and illicit content and contacts that teens and adolescents face on a near-daily basis.

- j. Responding to an article in the Guardian, “How Facebook and Instagram became marketplaces for child sex trafficking,” which was published in 2023, a Meta spokesperson promised: “The exploitation of children is a horrific crime – we don’t allow it and we work aggressively to fight it on and off our platforms.”¹³⁹

404. Meta’s Community Standards for Facebook declare: “We do not allow content or activity that sexually exploits or endangers children”¹⁴⁰ including child nudity, and, with respect to suicide/self-harm content proclaim: “We remove any content that encourages suicide or self-injury, including fictional content such as memes or illustrations and any self-injury content which is graphic, regardless of context.”¹⁴¹ Its Instagram Community Guidelines similarly claim: “We have zero tolerance when it comes to sharing sexual content involving minors or threatening to pose intimate images of others.”¹⁴² Meta’s Transparency Center promises that the company (in relevant part) “remove[s] content that facilitates or coordinates . . . human trafficking”, including “in order to force them to engage in commercial sex . . .” through “deception, force and coercion.”

¹³⁷ <https://about.instagram.com/blog/announcements/raising-the-standard-for-protecting-teens-and-supporting-parents-online>.

¹³⁸ <https://about.instagram.com/blog/announcements/continuing-to-keep-instagram-safe-and-secure>.

¹³⁹ <https://www.facebook.com/zuck/posts/10113961365418581>.

¹⁴⁰ <https://transparency.fb.com/policies/community-standards/child-sexual-exploitation-abuse-nudity/>

¹⁴¹ <https://transparency.fb.com/policies/community-standards/suicide-self-injury/>

¹⁴² <https://help.instagram.com/477434105621119>

However, as described above, such content is readily available on Facebook and Instagram and efforts to remove it are often rejected or unsuccessful.

405. Meta’s published “Community Standards” also proclaim that Meta “take[s] our role seriously in keeping abuse off the service.” But this statement cannot be squared with the numerous reports above of Meta failing to detect or address a huge volume of sexualized content, allowing children to access it, or designing its algorithms to recommend it and network it through a web of predators. Meta claims that such illicit or unlawful content is infrequent and is allowed to remain only after the company determines that its value outweighs its harm: “In some cases, we allow content—which would otherwise go against our standards—if it’s newsworthy and in the public interest. We do this only after weighing the public interest value against the risk of harm, and we look to international human rights standards to make these judgments.”¹⁴³

406. Meta’s efforts to publicly portray its platforms as safe and largely free of illicit content extends to quarterly Community Standards Enforcement Reports (“CSER”) which “provide metrics on how we enforced our policies . . . and estimates on the amount of violating content (Prevalence) on Facebook and Instagram.” Meta’s May 15, 2018 press release announcing the formation of these reports made clear that the reports were and are intended to allow the public to see “how much bad stuff is out there,” and thereby permit the public to “judge our performance for yourself.” Meta positioned itself as a company invested in eliminating illicit content from its platforms: “We believe that increased transparency tends to lead to increased accountability and responsibility over time, and publishing this information will push us to improve more quickly too. This is the same data we use to measure our progress internally – and you can now see it to judge our progress for yourselves.”

¹⁴³ <https://transparency.fb.com/policies/community-standards/> (last visited Dec. 1, 2023).

407. Each and every one of these reports underreport the existence of objectionable or violative conduct on Facebook or Instagram because they all rely on Meta’s flawed “prevalence” standard. A May 23, 2019 blog post described “prevalence” as “[o]ne of the most significant metrics we provide in the Community Standards Enforcement Report.” Meta reported that “we consider prevalence to be a critical metric because it helps us measure how violations impact people on Facebook. We care most about how often content that violates our standards is actually seen relative to the total amount of times any content is seen on Facebook.” It compared this metric to “measuring concentration of pollutants in the air we breathe” and claimed that “[p]revalence is the internet’s equivalent – a measurement of what percent of times someone sees something that is harmful.”¹⁴⁴

408. Meta’s CSERs consistently reported low prevalence of human trafficking, CSAM, bullying and other problematic materials. For example:

- a. The CSER released in November 2019 claimed that prevalence was an “upper limit [of] 0.04%” of views for content violating Meta’s policies prohibiting “child nudity and sexual exploitation of children, regulated goods, suicide and self-injury, and terrorist propaganda.”
- b. The December 2020 CSER claimed that “less than 0.05% of views were of content that violated our standards against Child Nudity and Sexual Exploitation” and that “less than 0.05% of views were of content that violated our standards against Suicide and Self-Injury.”
- c. The Q3 2021 CSER reported “that between 0.14% to 0.15% of views were of content that violated our standards against Bullying & Harassment” and that “less

¹⁴⁴ <https://about.fb.com/news/2019/05/measuring-prevalence/>.

than 0.05% of views were of content that violated our standards against Suicide & Self-Injury.”

409. Individually and collectively, each of these reports conveyed the impression that Meta aggressively enforced its Community Standards on both Facebook and Instagram, and that its efforts were succeeding in keeping the platforms relatively free of harmful content. For example, a November 13, 2019 news release announcing release of the fourth CSER includes the heading “Progress to Help Keep People Safe.” Meta’s most recent report (released in August 2023) claims that the purpose of the report is to “demonstrate our continued commitment to making Facebook and Instagram safe and inclusive.”

410. Nowhere do the CSERs explain how much sexualized content remains on the platforms and accessible to children; the ability of adult strangers to identify, groom, and seek sexualized content and activity from children; or the widespread sale of CSAM, among other commercial sexual exploitation of children.

411. Moreover, as explained above, the prevalence metric consistently underestimated the amount of problematic and illicit content displayed on Facebook. The prevalence metric contradicted the findings of Meta’s own BEEF study, which showed a much greater “prevalence” of bad experiences involving illicit, questionable or violative conduct on Meta’s platforms. The chart below demonstrates the discrepancies between the BEEF study (reported instances only over the prior 7 days) and the corresponding Community Standards Enforcement Report (which reports “prevalence” over a longer period of time):

<i>BEEF Study Metric</i>	<i>BEEF Study Finding (Overall)</i>	<i>Study Finding (Ages 13-15)</i>	<i>CSE Report (3Q 2021) Metric</i>	<i>CSE Report (3Q 2021)“Prevalence” on Instagram</i>
Witness to bullying (“bully witness”)	28.3%	27.2%	Bullying and Harassment	“between 0.05% to 0.06% of views”
Witness to hate message (“hate witness”)	25.3%	26.0%	Hate Speech	“about 0.02% of views”
Nudity	16.3%	19.2%	Adult Nudity & Sexual Activity	Displays range of 0.02% to 0.03%
			Child Nudity & Sexual Exploitation ¹⁴⁵	Not reported
“Unwanted Advances”	11.9%	13.0%	No corresponding category	
Target of bullying (“bully target”)	8.1%	10.8%	Bullying & Harassment	“between 0.05% to 0.06% of views”

412. Meta not only made affirmative misrepresentations, it also made material omissions in its public statements. For example, a June 8, 2021 blog post entitled “Shedding More Light on How Instagram Works” represented that Instagram employed “algorithms, classifiers, and processes” in order “to personalize your experience” to prioritize “what you care about most.”¹⁴⁶ But that post did not disclose that those recommendations could, and often did, include recurring posts containing illegal or harmful content, like that uncovered by the State’s investigation. Nor did it disclose the existence of “rabbit holes,” “vortices” or “flywheels,” well-known to Meta, that resulted from the algorithms design and operation. Further, an article in Instagram’s “Help Center”

¹⁴⁵ The CSER states that “Child Nudity & Sexual Exploitation” was “Measured until Q1 2021, and then replaced with Child Endangerment policy categories).” However, the CSER does not report prevalence for the “Child Endangerment: Sexual Exploitation” category or for the “Child Endangerment Nudity and Physical Abuse” category.

¹⁴⁶ <https://about.instagram.com/blog/announcements/shedding-more-light-on-how-instagram-works> (last visited Dec. 1, 2023).

section represents that Instagram “avoid[s] making recommendations that may be inappropriate for younger viewers.”¹⁴⁷

413. Many of Meta’s other public statements materially omitted facts known to Meta regarding the true nature of its platforms. While portraying its platforms as “safe,” Meta failed to disclose the internal studies in its possession—including the numerous internal studies cited above—identifying and quantifying the harm associated with using its platform, including studies related to self-image of teens and adolescents, studies related to causing anxiety and depression among teens and adolescents, and studies related to eating disorders and the prevalence of material related to eating disorders on Meta’s platforms.

414. Even though, as described above, Meta’s algorithms utilize artificial intelligence in order to “recommend” content or connections to users—including illicit and harmful content—Meta makes benign claims in documents such as its Instagram Privacy Policy that the data it uses the data it collects “[t]o promote safety, security and integrity,” “[t]o research and innovate for social good,” or “[t]o provide measurement, analytics and business services” for individuals who “rely on our Products to run or promote their businesses.”¹⁴⁸ In reality, Meta does not use its algorithms to identify illicit content. If it did, the results of the Attorney General’s investigation likely would not have occurred.

415. Further, Meta to this day continues to report its “prevalence” metric without disclosing the BEEF survey demonstrating the statistic materially misrepresented users’ experiences on its platforms.

¹⁴⁷ <https://help.instagram.com/313829416281232> (last visited Dec. 1, 2023).

¹⁴⁸ Privacy Policy, <https://privacycenter.instagram.com/policy>

416. All of these misrepresentations and omissions were made for purposes of pursuing profits over safety, and at the expense of teen and adolescent users of its platforms who suffered real and lasting harm due to Meta’s design and business choices.

XVIII. ADDITIONAL ALLEGATIONS RELATING TO MARK ZUCKERBERG

417. The documents and decisions described in this Complaint make clear that, on the decisions that mattered to children and parents in and beyond New Mexico, Mark Zuckerberg called the shots. Meta executives frequently raised with Zuckerberg issues regarding product features and design choices for “MZ [Mark Zuckerberg’s] feedback” or decision and, time after time, despite his professed commitment to protect children, Zuckerberg made decisions that caused children on Meta’s platforms to be less safe. He promised publicly that he directed that Meta safeguard children, but privately demanded and delivered the opposite. Zuckerberg, personally, was presented with requests for resources to ensure product integrity and safety, which he denied, and alarming statistics about the harm being caused by Meta’s products, including specific features, which he chose to continue or not address. Moreover, Zuckerberg, personally, made statements that misled the public about Meta’s conduct and its consequences, as laid out in this complaint.

418. From the company’s origin, Zuckerberg developed and defined the architecture of Facebook, and then Meta. He co-engineered the decisions to design Facebook to “consume as much . . . time and conscious attention as possible.”¹⁴⁹

419. Since then, Zuckerberg has knowingly misrepresented both his understanding and intention that Facebook was constructed to create addiction. Asked in a November 2020

¹⁴⁹ Alex Hern, *Never Get High on Your Own Supply—Why Social Media Bosses Don’t Use Social Media*, THE GUARDIAN (Jan. 23, 2018), <https://www.theguardian.com/media/2018/jan/23/never-get-high-on-your-own-supply-why-social-media-bosses-dont-use-social-media>.

Congressional hearing, “do you believe your product can be addictive,” Zuckerberg responded, “we certainly do not design the product in that way.”¹⁵⁰ Yet, the year before, on April 8, 2019, Meta’s Vice President of Product, Choice and Competition, David Ginsberg, emailed Zuckerberg with a request to “address problematic use” on Facebook and Instagram because internal and external research established that “problematic use” [i.e., addiction (which affected 58% of users), social comparison (45% of users), and loneliness (43% of users)] were the “three negative drivers that occur frequently on [our platform] and impact people’s well-being.” A few months later, on August 30, 2019, still before his 2020 appearance before Congress, Facebook staff emailed Zuckerberg to prepare him for a meeting with a “leading researcher on the topic of social media and its effect on mental health and suicide amongst teens.” While his staff voiced skepticism of the causal link, they counseled Zuckerberg to acknowledge that “we understand that Facebook can be used in problematic ways and are proactively investing in research, product improvements, and campaigns, to address these issues.”

420. In testimony to Congress in March 2021, Zuckerberg falsely denied that he made “money off of creating an addiction to [his] platforms”¹⁵¹ and contended that he did not set “goals around increasing the amount of time that people spend” on the platform.¹⁵² As laid out above, Meta has developed, implemented, and declined to rein in product features—such as autoplay, ephemeral content, infinite scroll, and notifications— that were designed to hook users, especially teenage users, on Facebook and Instagram. Upon information and belief, Meta’s “Teen Mental

¹⁵⁰ *Breaking the News: Censorship, Suppression, and the 2020 Election: Hearing Before the S. Comm. on the Jud.*, 116 Cong. (2020) (Statement of Mark Zuckerberg); see also *Facebook CEO Mark Zuckerberg on Whether Products are Addictive: “We Certainly Do Not Design the Product in that Way[,]”* THE RECOUNT, (Nov. 17, 2020), <https://therecount.com/watch/facebook-ceo-mark-zuckerberg-on/2645864077>.

¹⁵¹ *Disinformation Nation: Social Media’s Role in Promoting Extremism and Misinformation: Hearings Before the H. Subcomms. on Comm’n & Tech., Consumer Prot. & Commerce, and Comm. on Energy & Com.*, 117 Cong. at 107:2491-2497 (2021) (Statement of Mark Zuckerberg), <https://www.congress.gov/117/meeting/house/111407/documents/HHRG-117-IF16-Transcript-20210325.pdf>. (last visited Dec. 1, 2023).

¹⁵² *Disinformation Nation*, *supra* note 157, at 69:1551-1575

Health Deep Dive” research confirmed to Zuckerberg that Meta’s design was working, instilling in teenagers “an addicts’ narrative about their use” who “wish they could spend less time caring about it” but they “can’t help themselves.”

421. Contrary to his testimony, Meta—and Zuckerberg specifically—have explicitly aimed to increase the time spent on the platform, especially by teens. In a December 2015 email to Susan Li, Meta’s Chief Financial Officer, and David Wehner, the Chief Strategy Officer, Zuckerberg described his goals for 2016, including to see “[t]ime spent increase[] by 12%” within three years and, for Instagram, by 10% within five years. A Facebook presentation titled “2017 Teens Strategic Focus” described Meta’s goal, consistent with “Mark’s [decision] that the top priority for the company in 2017 is teens” to “increase U.S. teen time spent” through Instagram direct and stories.

422. Zuckerberg also deceptively portrayed Facebook as a safe space for children, despite substantial evidence of the harms to children through both the child exploitation and mental health consequences described in this Complaint. In a message he posted on October 6, 2021, Zuckerberg sought to reassure that “We care deeply about issues like safety, well-being and mental health,” at the same time that Zuckerberg was making decisions not to address harmful features on the platform like cosmetic filter or address findings on addiction, self-harm, and unwanted sexual advances to teenagers. At the 2021 Congressional hearing, Zuckerberg stated that he did not believe that Facebook harms children, explaining, “This is something that we study and we care a lot about; designing products that [sic] peoples’ well-being is very important to us.”¹⁵³

423. Zuckerberg doubled down in follow-up testimony to Congress after Frances Haugen’s disclosure of Facebook documents, and a post on his Facebook page, assuring consumers

¹⁵³ *Disinformation Nation*, *supra* note 157, at 175: 4166–4175.

personally that “it’s very important to me that everything we build is safe and good for kids.” He continued: “when it comes to young people’s health or well-being, every negative experience matters. It is incredibly sad to think of a young person in a moment of distress who, instead of being comforted, has their experience made worse.”¹⁵⁴

424. In order to maintain this public image, Zuckerberg and his team crafted his public comments to omit information that revealed Meta’s deficiencies in addressing child sexual exploitation on its platforms. In a May 20, 2019 email, Derick Mains, a communications staffer, briefed Zuckerberg in advance of a CSER call and provided a script for his public comments. While the background provided to Zuckerberg listed “child nudity and sexual exploitation of children, along with bullying and harassment and adult nudity and sexual activity, as ‘Areas of regression,’” Zuckerberg’s comments listed only “bullying, harassment and nudity” as issues on which “we have a ways to go,” omitting the critically important and identified deficiencies regarding child sexual exploitation.

425. Inconsistent with his public representations regarding Meta’s commitment to child safety, Zuckerberg had been specifically warned that Meta’s policies with regard to teen suicide and self-harm were inadequate and indefensible. A January 26, 2019 email thread involving Nick Clegg, Vice President of Global Affairs, and Antigone Davis began with a comment from Chris Norton:

As you know, the Sunday Times will tomorrow run a (likely) front-page story with 30 families of suicide victims accusing Instagram of killing their children. The Government will write to us to tell us we need to do more and that this is a watershed moment. In my view it is. The Instagram team have been handling and doing their best to communicate why we have the current set of policies. I understand that there is an XFN [(cross functional team)] that is looking at those policies and trying to get experts to defend us. We

¹⁵⁴ Salvador Rodriguez, *Zuckerberg Rejects Claims That Facebook Prioritizes Profits Over User Safety*, CNBC (Oct. 5, 2021), <https://www.cnbc.com/2021/10/05/zuckerberg-denies-that-facebook-prioritizes-profits-over-user-safety.html> (last visited Dec. 1, 2023).

are defending the status quo when the status quo is clearly unacceptable to media, many impacted families, and when revealed in press, will be unacceptable to the wider public.

Despite having been warned that Meta’s policies with respect teen self-harm and to age limits and restrictions for content were “unenforced” and insufficient, Zuckerberg both failed to take appropriate steps to protect children from harm on the platforms, failed to publicly disclose the risks of self-harm and suicide associated with use of Meta’s platforms, and continued to misrepresent that Meta’s platforms are “safe and good” for kids.

426. Contrary to his promised commitment to safeguard children and other users of the platforms, Zuckerberg over and over again made or approved decisions that put user engagement and Meta’s revenue ahead of its users’ health and safety. On April 14, 2017, Instagram co-founder Kevin Systrom emailed Zuckerberg to ask for more staff for “Integrity” at Instagram. Zuckerberg noted that he had committed to additional staff for integrity and promised to include “IG in this mix,” but pointed out that Facebook had “more extreme issues . . . with the [Christchurch] murder, bad activity in private groups, etc” – making clear that Meta still had not addressed critical safety issues on its legacy platform.

427. Fearing that Zuckerberg did not “understand the urgency of working on integrity related issues at IG,” Systrom reached out to colleagues to collect examples of equally dangerous issues at Instagram, which included that “46% of [users] surveyed reported experiencing bullying/harassment on IG” and that “30% [of those users] were harassed 4-10+ times”; “50% reported receiving unwanted or offensive comments”; “[t]eens delete twice the [percentage] of comments from unconnected commenters”; “25% of all deleted comments come from repeat bullies”; and “[m]any of the profiles that are ultimately disabled have not posted anything and appear to be created expressly to troll comment threads.” The list also noted that there was no

detection system for self-harm and bullying and that: “recently, ‘death group’ hashtags in Russia encouraged teenagers to commit suicide. We don’t have a systematic way to discover or monitor these hashtags” except manually. “Shielded accounts” have livestreamed sex acts, but, because of their status, were “immune” to attempts to take down their posts.

428. With this backdrop, Systrom replied to Zuckerberg on April 21, 2017 to note that issues were equally bad at Instagram, where “[l]ast weekend a boy shot himself on IG live” and there were an average of three live videos a week involving “imminent danger from issues ranging from self harm to child exploitation.” His email also confirmed that celebrities had remedies available to them that were not available to average teen users. Bullying of some of Instagram’s top accounts, such as Justin Bieber, Beyonce, and Selena Gomez, got special attention or resulted in the accounts being deleted. But the typical teen user did not receive similar treatment from Instagram and, as Meta already knew, could not readily delete their accounts.

429. Likewise, Ginsberg’s April 2019 email to Zuckerberg, described above, also noted the need for additional engineering staff focused on building well-being “tools/products to address problematic use” on the platform because “[c]urrent research (internal and external) tells us that . . . there is increasing scientific evidence (particularly in the US. . .) that the average net effect of [our platforms] on people’s well-being” is “slightly negative.” The email asked for 24 additional staff to understand problematic use on Instagram and to “to build tools/products to address problematic use” on Facebook. The email warned that, without the investment, research would proceed at “a slower pace (and NO product changes).” Particularly at Instagram, Ginsberg warned that there would be no other staff to devote to the work. CFO Susan Li, presumably communicating Zuckerberg’s decision on this issue, responded to the email and rejected the requested staffing, though noting that “both Mark and Sheryl wanted to emphasize they think this is work worth

doing.” Even though Facebook’s revenue was more than \$70 billion, an increase of more than \$15 billion over 2018,¹⁵⁵ Li warned that any additional staff would be hard to come by, and Mosseri responded that Instagram would not fund it either.

430. A few months later, in September 2019, the heads of Instagram and Facebook, Mosseri and Fidji Simo, respectively, emailed regarding how to address users’ well-being. Mosseri recounted the history:

Early 2019, Chris Cox was looking into us doing more on [well-being], came to David Ginsberg and asked him what it would take to 10x well-being, which resulted in a plan that was reviewed with Chris Cox and Sheryl Sanberg and then was submitted to Mark as an incremental ask, maybe -10 engineering and -15 XFN. They were told to do this, but self-fund, which ended up punting given other ambiguities.

Zuckerberg, in other words, did not make additional staff available to scale up Meta’s work to improve well-being. In defining the problem, they “100% agree[d]” that “the main problem is that we need to increase investment.” Mosseri observed that “Well-being is the existential question we face,” and contrary to Zuckerberg’s representations, Meta “lack[ed] . . . a roadmap of work that demonstrates we care about well-being.”

431. More than four years after Systrom initially sounded the alarm to Zuckerberg, on August 21, 2021, members of Instagram’s Well-being Team sent an email to senior executives laying out the need to add staff to the “currently underinvested” but critical areas of “problematic use,” as well as “bullying+harassment, connections, [and Suicide and Self-Injury (SSI)]” for teens, noting that “[t]hese topics are highly aligned with what teens want Facebook and Instagram to prioritize.” On August 27, 2021, Clegg forwarded the email to Zuckerberg, endorsing the request and saying it was “increasingly urgent” to address “concerns about the impact of our products on

¹⁵⁵ https://s21.q4cdn.com/399680738/files/doc_news/Facebook-Reports-Fourth-Quarter-and-Full-Year-2019-Results-2020.pdf

young people’s mental health.” Clegg explained that “we need to do more and we are being held back by a lack of investment on the product side which means that we’re not able to make changes and innovations at the pace required.” He said the company’s “wellbeing work is both understaffed and fragmented” and warned that “[w]e are not on track to succeed for our core well-being topics (problematic use, bullying & harassment, connections, and SSI).”

432. Despite Zuckerberg’s statements that he was personally committed to ensuring that Meta’s products were “safe and good for kids” and the urgency of these issues for both teens’ mental health and Meta’s public image, Zuckerberg did not respond to Clegg’s email. Three months later, on November 10, 2021, Clegg sent Zuckerberg a follow-up email with a “scaled . . . back” request on the number of staff, but urged that “this investment is important to ensure we have the product roadmaps necessary to stand behind our external narrative of well-being on our apps.” Susan Li again responded for Zuckerberg to note “constrained” staffing.

433. Instead of approving the requested staffing needed to meet Meta’s avowed commitment to teens’ well-being, upon information and belief, Zuckerberg would have been the one to have approved, if not initiated, the decision in September 2022 to disband Meta’s Responsible Innovation Team and cut at least 16 members of Instagram’s well-being group and more than 100 positions related to trust, integrity and responsibility.¹⁵⁶

434. In the same spirit of disregard for teen’s mental health and safety, Zuckerberg decided to retain cosmetic surgery filters despite an internal consensus that these filters harm teenagers, especially teenage girls, out of concern that this step might decrease engagement on the

¹⁵⁶ Jeff Horwitz, *Facebook Parent Meta Platforms Cuts Responsible Innovation Team*, WALL ST. J. (Sept. 8, 2022), <https://www.wsj.com/articles/facebook-parent-meta-platforms-cuts-responsible-innovation-team-11662658423>; Brandon Vigliarolo, *Meta Disbands Responsible Innovation Team, Spreads it out Over Facebook and Co*, THE REGISTER (Sept. 9, 2022), https://www.theregister.com/2022/09/09/meta_disbands_responsible_innovation_team/; Hayden Field, et al., *Tech Layoffs Ravage the Teams that Fight Online Misinformation and Hate Speech*, CNBC (May 26, 2023), <https://www.cnbc.com/2023/05/26/tech-companies-are-laying-off-their-ethics-and-safety-teams-.html>.

platforms. Cosmetic surgery filters allow users to manipulate photos to model the impacts of cosmetic surgeries or other cosmetic “enhancements” and expose users to photos filtered by others.¹⁵⁷

435. As Margaret Gould Stewart, Vice President of Product Design, recounted after a “PR fire” in October 2019, Meta put in place a temporary ban on cosmetic filters. On November 12, 2019, Gould Stewart emailed Mosseri, Simo, and the Head of Policy for Instagram, Karina Newton, to seek support to bar “effects that mimic plastic surgery,” including plastic surgery filters because of the negative “impacts that th[e]se effects were having on mental health and wellbeing, especially for more vulnerable users (youth, women).” There was broad support for Gould Stewart’s proposal until Andrew Bosworth, Chief Technology Officer, offered that Zuckerberg “might want to review before implementing” because Zuckerberg questioned whether these filters actually “represent[] real harm.”

436. Gould Stewart indicated that she was “surprised” by Zuckerberg’s reaction since “all teams were in support of these guidelines” and provided “academic research and external engagement that was conducted before forming the policy recommendations” that “outlines some compelling motivations for formalizing these policies.” Newton also responded to support removing the filter, which is “overwhelmingly used by teen girls”: “it’s been our strong recommendation from comms, marketing, policy, and engagement with nearly 20 outside experts and academics that we pass this policy.” She continued: “we’re talking about actively encouraging young girls into body dysmorphia [a mental health condition involving a focus on perceived flaws in appearance] . . . that can result in serious issues.” Background materials also

¹⁵⁷ Help Center, *Apply Filters to Your Post on Instagram*, INSTAGRAM, <https://help.instagram.com/453965678013216> (last visited 11/14/2023)

noted that Google had invested heavily in research on “camera beautification” and “their conclusions have been similar to ours.”

437. Zuckerberg received a briefing (or “pre-read”) in advance of a meeting set for April 2, 2020 “to decide whether to continue, modify or lift the temp[orary] ban on Cosmetic Surgery ARR that [they] instituted in October 2019.” The pre-read repeated the consensus of nearly two dozen independent experts consulted by Meta that “these effects are cause for concern for mental health and wellbeing, especially amongst vulnerable populations (females, youth, those with a history of mental health concerns, etc.)” and that, especially for children, “[t]hese extreme cosmetic surgery effects can have severe impacts on both the individuals using the effects and those viewing the images.” The meeting was abruptly cancelled and Zuckerberg instead emailed his desire to find “a way to relax or lift the ban on cosmetic effects,” which Zuckerberg then formalized. The pre-read briefing clearly stated that this option “still [had] notable wellbeing risk (since recs [were] a marginal source of virality).” Saying, despite the pre-read’s detailed summary of the evidence, that he had seen “no data” that filters were harmful, Zuckerberg cited “clear[] demand” for the filters.

438. Gould Stewart warned Zuckerberg: “I don’t think it’s the right call given the risks. ... I just hope that years from now we will look back and feel good about the decision we made here.”

439. In an April 2020 email, Zuckerberg provided his personal views, which contradicted much of the internal evidence Meta had gathered. Notwithstanding Meta’s “engagement with nearly 20 outside experts,” Zuckerberg remained skeptical: “It has always felt paternalistic to me that we’ve limited people’s ability to present themselves in these ways, especially when there’s no data I’ve seen that suggests doing so is helpful or not doing so is

harmful, and that there's clearly demand for this type of expression. I'm generally supportive of finding a way to relax or lift these bans." Zuckerberg ultimately supporting "lifting restrictions" on the cosmetic surgery filters "but not promoting on IG." He made this recommendation even though the supporting documentation noted that "[r]eversing this decision undermines Instagram's well-being commitment and jeopardizes any future efforts to work with policymakers and mental health experts on these issues."

440. Similarly, on April 22, 2020, "MZ Feedback" was shared with the Integrity XFN. Zuckerberg's feedback addressed recommendations of the Soft Actions Task Force, which was tagged to provide "Big ideas to reduce prevalence of bad content in News Feed." The Task Force noted that, in addition to removing violating content (a hard action), Facebook could take soft actions, short of removal, such as demoting the placement of content or users in the platforms' feeds, that would "reduce the prevalence of bad content," such as hate, graphic violence and, relevant here, nudity or pornography (or "N&P"), effectively burying this content so that users were less likely to encounter it. The Task Force estimated that Meta could remove roughly 17% of N&P (with a margin of +/- 10%) and could reduce the prevalence of N&P by another roughly 17%--doubling its efficacy--by demoting content that was *probably* violating nudity and pornography standards, and another 17% by demoting *Probable* Violating Accounts. The Task Force expected that these demotions would not result in a "decrease in likes," interactions, or sessions. However, Zuckerberg would only support demoting Probable Violating Content as a "break the glass measure" and was "[n]ot supporting of account-level demotion, so we will NOT be launching this." When an employee comment in an internal message board asked whether this applied also to terrorism, drug sales, and human trafficking, the response was "Generally yes." Because of the risk that some content might be wrongly flagged, Zuckerberg, dangerously and

contrary to his public assurances, chose the path that made Facebook less “safe” and less “good” for kids--and everyone else.¹⁵⁸

441. Zuckerberg is also personally responsible for his failure to take action to address, and Facebook’s repeated misrepresentations regarding, users’ negative experiences on Facebook and Instagram. On October 5, 2021, Bejar emailed Zuckerberg and other senior executives regarding a “critical gap” between Meta’s public statements regarding the prevalence of harmful content on its platform and the actual experiences of users. Bejar summarized and attached the results of the BEEF survey that provided “ground-level truth” regarding the impact that Facebook and Instagram were having on its users, especially young users. This should have been a five-alarm call to Zuckerberg, but he—once again—failed to respond.

442. The email provided a dark picture of the platform’s success in addressing harmful content: 51% of Instagram users said “yes” to having had a bad or harmful experience in the last 7 days. Nearly 22% of 13-15 year olds were the targets of bullying on the platforms and 39% of 13-15 year olds experienced negative comparisons, Bejar reported. Most shocking, “24.4% of 13-15year old[s] . . . said they received unwanted advances.”

443. Bejar asked Zuckerberg: “what if policy based solutions,” grounded in publicly reported prevalence data, “only cover a single digit percentage of what is harming people?” Bejar urged Zuckerberg to prioritize and fund initiatives that would identify the content that is causing bad experiences for users, especially the most intense bad experiences, and to determine how much of that content violates Meta’s current policies – which is what drives (and, thus, what limits) the identification, removal, and reporting of illegal, negative, and dangerous content and ensures that it is not amplified through Meta’s algorithms. Zuckerberg, who Bejar stated typically responded

¹⁵⁸ https://s3.documentcloud.org/documents/23596616/tier3_force_ro_0420.pdf

to his emails, did not respond to him. Two years later, Meta and Zuckerberg continue to publicly disclose and rely on the prevalence of policy-violating content to promote the success of its efforts to address activity that degrades teens' mental health and sexually exploits children, rather than devoting more (or even maintaining current) resources to understand and address this content.

444. Deposed during attorneys general's investigation of Meta, Bejar testified that he believed that Zuckerberg and other company leaders focused on a prevalence measure "because it created a distorted picture about the safety of Meta's platforms." When asked if he thought "Mr. Zuckerberg's public statements about prevalence created a misleading picture of the harmfulness of Meta's platforms," Bejar testified, "I do." Bejar made this point during his testimony to attorneys general, saying that each time prevalence statistics are used publicly, "they're minimizing the harms that people are experiencing in the product." Creating both an image and strategy built on the falsehood that children and teens were devotedly and fully protected on the platforms, Zuckerberg not only deceived the public but undermined Meta's ability to obtain accurate information that would have better enabled it to identify, remove, and avoid amplifying features and content that harm children and teens.

445. Zuckerberg failed to ensure that Meta corrected its algorithms for recommending users and content, despite the company's knowledge that the algorithms were providing "egregious content" to young users. Meta did not adopt design changes it tested, even though they produced "a meaningful drop in exposure," because such changes "came with a clear engagement cost."

446. In the same vein, in 2018, Zuckerberg also made clear that the number of users on Facebook ("DAP," or daily average persons) would take priority over any changes to Meta's policies to reduce notifications. As one employee noted: "Fundamentally, I believe that we have abused the notifications channel as a company." Cox agreed and offered that Meta should not

back off from doing what was “better for people” because usage metrics were down. Zuckerberg disagreed. “US DAP is a bigger concern for Mark right now than user experience.” The email continued: “I understand that NF/Notifications serve many interests, but we got a very clear and strong message from Mark that DAP (and specifically US DAP) is extremely important and we must change the trajectory from the negative one it’s on now.”

447. Zuckerberg also personally made decisions on whether posts or content should be removed as violating Meta policy and, by implication, made decisions on whether posts or content would remain on Meta’s platforms. Internal documents provided by Haugen acknowledged that “Facebook routinely makes decisions about algorithms based on input from Public Policy.” Unlike the team devoted to Content Integrity, Public Policy “typically are interested in the impact on politicians and political media, and they commonly veto launches [product changes] which have significant negative impacts on politically sensitive actors.” Further, “Final calls about content policy are routinely made by senior executives,” and “In multiple cases the final judgment about whether a prominent post violates a certain written policy are made by senior executives, sometimes Mark Zuckerberg.” The author(s) then questioned: If our decisions are intended to be an application of a written policy then it’s unclear why executives would be consulted.”

448. Upon information and belief, given that public attention to Meta’s well-being and child safety initiatives would have elevated the importance of the issues and Zuckerberg’s involvement in other similar decisions, Zuckerberg was responsible for making, or declining to reverse, numerous other decisions to drop or limit proposals that would have decreased teens’ addiction or mental health harms and protected children from child sex exploitation and abuse. These included deploying an ineffective “time spent” tool that fails to help teenagers limit their use of the platforms, rejecting a default setting to hide “likes,” and refusing to turn off notifications

to teenagers at night. Zuckerberg also would have been the decision-maker for the termination of research and surveys that revealed negative information and for directing Meta employees to not document critical findings and proposals to address these findings, along with drastic cuts to the number of human investigators and other safety, well-being, and integrity personnel.

449. The fact that Zuckerberg acted within the broad scope of his position as Meta’s founder, chair, and CEO, is not a defense to his liability; it is, instead, an indictment of his failures. However, it is because of the role that Zuckerberg, individually, played in directing, carrying out, ratifying, and refusing to acknowledge or address Meta’s unlawful, unfair and deceptive conduct that Zuckerberg is individually liable for the violations of law described in this Complaint.

XIX. CAUSES OF ACTION

COUNT I VIOLATION OF NEW MEXICO UNFAIR PRACTICES ACT (UNFAIR OR DECEPTIVE TRADE PRACTICES) NMSA 1978, § 57-12-1 to -26 (Against All Defendants)

450. The State re-alleges all prior paragraphs of this Complaint as if fully set forth herein.

451. The Unfair Practices Act (“UPA”) prohibits “[u]nfair or deceptive trade practices and unconscionable trade practices in the conduct of any trade or commerce.” NMSA 1978, § 57-12-3 (1971).

452. Defendants are engaged in “trade” or “commerce” as defined by the UPA, which “includes the advertising, offering for sale or distribution of any services and any property and any other article, commodity or thing of value, including any trade or commerce directly or indirectly affecting the people of this state.” NMSA 1978, § 57-12-2(C) (2019). Defendants advertise, offer and distribute their internet platforms within New Mexico and to New Mexico residents.

453. The UPA defines an “unfair or deceptive trade practice” as “a false or misleading oral or written statement, visual description, or other representation of any kind knowingly made in connection with the sale . . . of goods and services . . . in the regular course of the person’s trade or commerce, that may, tends to or does deceive or mislead any person.” Section 57-12-2(D).

454. The UPA provides an inclusive rather than exhaustive list of examples of unfair or deceptive trade practices. These include the following: “representing that goods or services are of a particular standard, quality or grade . . . if they are of another[;]” “making false or misleading statements of fact concerning the price of goods or services[;]” and “using exaggeration, innuendo or ambiguity as to a material fact or failing to state a material fact if doing so deceives or tends to deceive[.]” Section 57-12-2(D).

455. Meta does not require a monetary exchange from New Mexican consumers in order for them to use Meta platforms, but Meta made representations about its platforms in connection with the sale of goods and services. Specifically, Meta sells advertising, and Meta increases its ad revenue by requiring consumers to agree to the use of their private data for targeted advertising. Meta collects its users’ data and then uses it to generate revenue.

456. In addition to offering, advertising, and distributing its social media platforms in New Mexico, Meta thus receives revenue both for showing ads to New Mexico consumers and also for harvesting New Mexico consumers’ personal data, including information about their activities and interests, to target advertising and thereby increase its revenue from selling ads. Meta’s platforms also facilitate the sale of goods and services, both through advertisements that Meta directs to New Mexico residents and within New Mexico in exchange for a fee, and by providing space for users to offer, buy and sell merchandise (“Facebook Marketplace”). Meta charges and collects a fee when items are sold on its Marketplace, including in New Mexico. Meta

enables users to monetize their accounts in order to sell subscriptions or permit advertisements to be placed on their platforms. Additionally, users may receive “stars” that other users purchase from Meta and those “stars” can be monetized by the user as well.

457. At all times relevant herein, the Defendants violated the UPA, NMSA 1978, §§ 57-12-1 to -26 (1967, as amended through 2019), by committing repeated and willful unfair or deceptive acts or practices in the conduct of commerce, both of which are violations of the UPA.

458. The Attorney General is authorized to bring an action in the name of the State to remedy violations of the UPA. NMSA 1978, §§ 57-12-8(A) (1978), 57-12-15 (1967). This action is proper in this Court because Defendants are using, have used, and continue to use practices that are unlawful under the UPA. Section 57-12-8(A).

459. Defendants’ conduct, as alleged in the foregoing paragraphs, violated the UPA because Defendants knowingly made numerous false or misleading oral or written statements, visual descriptions, or other representations in connection with the sale of goods and services that had the capacity or tendency, or actually did, deceive or mislead any person.

460. In numerous instances, Defendants’ public statements and communications knowingly misrepresented, directly or indirectly, expressly or by implication, that their platforms were not addictive, that they prioritized young users’ well-being over profits, and that their platforms were safe, while concealing and/or misrepresenting their internal knowledge that the frequency of harms and harmful material or conduct encountered by young users on their platforms was far more pervasive than Defendants’ public statements revealed.

461. Specifically, Defendants have willfully, knowingly, and repeatedly violated the Unfair Practices Act by engaging in multiple deceptive acts and practices that duped young users, their families, and the public regarding the safety of their platforms and Meta’s efforts in

prioritizing well-being. Defendants engaged in misrepresentations, omissions, and/or active concealment to news media and the general public, including New Mexico children, that falsely and misleadingly asserted that:

- a. Users of Defendants' social media platforms did not encounter significant levels of harmful content, including content related to CSAM and human trafficking, including by relying on a misleading prevalence metric that Defendants had reason to know did not accurately reflect the levels of harmful content appearing on their platforms;
- b. Defendants would take actions to enhance the safety of their platform even at the expense of profits or losing users when, in fact, Defendants would subordinate safety decisions to other factors, including profits or competitive advantage;
- c. Defendants dedicated sufficient and adequate resources to policing its social media platforms (including removal of banned content) when, in fact, internal documents confirmed that Defendants' resources were admittedly inadequate to address the vast amounts of harmful, illicit, and inappropriate content appearing on its platforms;
- d. Defendants reported all illicit CSAM material on their platforms to the proper authorities, including to the NCMEC, when, in fact, significant numbers of illicit CSAM material on Defendants' platforms were not reported;
- e. Defendants' social media platforms, including Facebook and Instagram, are not designed to be addictive when they are so designed;

- f. Defendants prioritized young users' well-being, when in fact Defendants repeatedly chose not to invest in well-being initiatives and deliberately decided not to implement measures they knew could reduce harms to youth;
- g. Defendants' research regarding the effect of removing or hiding public "like" counts on its content (Project Daisy) was inconclusive, when in fact the research demonstrated that such an action would have significantly improved users' well-being;
- h. Defendants' algorithms are designed to "tailor" an experience to a user, when in fact the algorithms are designed to increase usage and engagement on the Meta's platforms;
- i. Defendants collect data for benign purposes, such as "to research and innovate for social good," or "to provide measurement, analytics and business services" for third parties, or for individuals who "rely on our Products to run or promote their businesses," when in fact Defendants collect data in order to ensure users remain on their platform and in order to increase engagement with Meta's platforms at all costs;
- j. Defendants' social media platforms, including Instagram, are safe for young users while concealing their internal research showing the high frequency at which young users experienced harms from their use of the platforms or viewed content or encountered activities on its platforms that Defendants had identified as harmful; and
- k. Defendants' protocols for preventing use of their platforms by children under the age of 13 were adequate when, in fact, Defendants internally

admitted their measures were little more than pretext and that children readily lied about their age in order to gain access to certain features of the platforms.

462. Defendants failed to disclose the harmful effects, content, and activities on its platforms, including as laid out below, which was misleading particularly in light of affirmative statements regarding the safety of their platforms:

- a. Failed to disclose the incidence and risk of addiction, depression, anxiety, sleep deprivation, eating disorders, suicide, negative self-image and dysmorphia, and other self-harms associated with use of its platforms by young users;
- b. Failed to disclose the incidence and risk of exposure to CSAM, sexually explicit and other inappropriate activity and content by young users, and the grooming, solicitation, and sexual advances of young users of its platforms;
- c. Failed to disclose that Defendants failed to adequately address the existence of sexually explicit conduct and activity and other self-harm or harmful content on its platforms and connected young users to such dangerous content and users through its algorithms;
- d. Failed to disclose that its algorithms were designed to leverage young users' dopamine responses and create an addictive cycle of engagement;
- e. Failed to disclose that its algorithms collect data in order to fuel young users' compulsive use of Meta's platforms;

- f. Failed to disclose that its algorithms will deliver harmful content to a user who clicks on such content, and failed to disclose the existence of “flywheels” or “rabbit holes” frequently caused by Meta’s algorithms;
- g. Failed to disclose its knowledge that the use of its prevalence data misrepresented the scope and severity of harms to young users of its platforms;
- h. Failed to disclose its knowledge that certain features of its platforms, including, but not limited to, its algorithms, the presence of public “like” counts, the use of “infinite” or “ephemeral” content and its inclusion of image filters, had detrimental effects on the well-being of young users;
- i. Failed to disclose that Meta did not remove all banned content it encountered on its platforms;
- j. Failed to disclose that Meta knew it had, and continued to establish, user accounts for children under 13 years of age, failed to screen those accounts from inappropriate and unlawful activity and conduct, and collected and used data from those children; and
- k. Failed to disclose that Meta continued features, such as likes, notifications, recommended posts, groups, and accounts, and filters, despite knowledge of the harms these features posed to young users.

463. Specifically, and incorporating the allegations above, Zuckerberg personally made statements or omissions that misrepresented: (i) the safety of Meta’s platforms for children; (ii) Meta’s knowledge of the harms to children from using its platforms, including addiction, suicide, and other self-harm and mental health consequences; (iii) his and Meta’s decision-making to design Meta’s platforms to be addictive and to increase users’ time on the platforms; (iv) the

frequency with which children experienced harms on Meta's platforms, and Meta's knowledge regarding the scope and severity of these harms; (v) his and Meta's commitment to make decisions to ensure the safety of children on Meta's platforms, when Zuckerberg and the company acted to disregard their safety and mental health, and (vi) the adequacy and efficacy of Meta's efforts to address CSEC on its platforms and its knowledge that predatory content and activities continued to proliferate.

464. These statements and omissions were made to falsely reassure young users, their parents, and the public that Meta's platforms were safe so that Meta could continue to attract, retain, and engage young users and thereby increase Meta's revenues, and have prevented consumers in New Mexico from taking steps to protect their health and well-being.

465. These statements and omissions were deceptive and misleading in that they, *inter alia*, conveyed a false impression that Meta's platforms had characteristics and benefits that they did not; represented that Meta's platforms were of a particular standard, quality or grade that they were not; and exaggerated, omitted, and created ambiguity as to facts that Meta recognized were material and that deceived or tended to deceive consumers.

466. Each deceptive act or practice engaged in by Defendants as recited above and throughout this Complaint constitutes a separate violation of the Unfair Practices Act.

467. New Mexico consumers and youth are suffering, have suffered, and will continue to suffer unjustified substantial injury as a result of Defendants' violations of New Mexico laws. Absent injunctive relief by this Court, Defendants are likely to continue to injure consumers and harm the public interest.

468. Plaintiff, State of New Mexico, seeks all legal and equitable relief as allowed by law, including *inter alia* injunctive relief, disgorgement of profits, all recoverable penalties under

Section 57-12-11 including a civil penalty of \$5,000 per each violation per each Defendant named in this Count, attorney fees and costs, and pre- and post-judgment interest.

COUNT II
VIOLATION OF NEW MEXICO UNFAIR PRACTICES ACT
(UNFAIR TRADE PRACTICES)
NMSA 1978, § 57-12-1 to -26
(Against All Defendants)

469. The State re-alleges all prior paragraphs of this Complaint as if fully set forth herein.

470. The UPA prohibits “[u]nfair or deceptive trade practices and unconscionable trade practices in the conduct of any trade or commerce.” NMSA 1978, § 57-12-3 (1971).

471. Defendants are engaged in “trade” or “commerce” as defined by the UPA, which “includes the advertising, offering for sale or distribution of any services and any property and any other article, commodity or thing of value, including any trade or commerce directly or indirectly affecting the people of this state.” NMSA 1978, § 57-12-2(C) (2019). Defendants advertise, offer and distribute their internet platforms within New Mexico and to New Mexico residents.

472. In addition to offering, advertising, and distributing its social media platforms in New Mexico, Meta receives revenue both for showing ads to New Mexico consumers and also for harvesting New Mexican consumers’ personal data, including information about their activities and interests, to target advertising, thereby increasing its revenue from selling ads. Meta’s platforms also facilitate the sale of goods and services, both through advertisements that Meta directs to New Mexico residents and within New Mexico in exchange for a fee, and by providing space for users to offer, buy and sell merchandise (“Facebook Marketplace”). Meta charges and collects a fee when items are sold on its Marketplace, including in New Mexico. Meta enables users to monetize their accounts in order to sell subscriptions or permit advertisements to be placed

on their platforms. Additionally, users may receive “stars” that other users purchase from Meta and those “stars” can be monetized by the user as well.

473. At all times relevant herein, the Defendants violated the UPA, NMSA 1978, §§ 57-12-1 to -26 (1967, as amended through 2019), by committing repeated and willful unfair or deceptive acts or practices in the conduct of commerce, both of which are violations of the Act.

474. The Attorney General is authorized to bring an action in the name of the State to remedy violations of the UPA. NMSA 1978, §§ 57-12-8(A) (1978), -15 (1967). This action is proper in this Court because Defendants are using, have used, and continue to use practices that are unlawful under the UPA. Section 57-12-8(A).

475. Defendants’ conduct, as alleged in the foregoing paragraphs, constitutes unconscionable trade practices because their acts and practices: (1) take advantage of the lack of knowledge, ability, experience or capacity of New Mexico consumers—especially children—to a grossly unfair degree; and (2) results in a gross disparity between the value received by the consumer and the price paid. Moreover, Defendants engaged in unfair acts and/or practices within the meaning of the UPA because their acts and practices are: (1) offensive to public policy as reflected in common-law, statutory, or other established expression of public policy; (2) immoral, unethical, oppressive, unscrupulous, and unconscionable; and/or (3) have caused unjustified, substantial injury to consumers that consumers themselves could not reasonably have avoided.

476. Defendants’ conduct as described above constitutes unfair acts and/or practices within the meaning of the Unfair Practices Act because, as explained above, Defendants’ acts and practices are coercive, exploitative, abusive, deceptive, and/or predatory. The conduct described above involves the intentional manipulation of youth behavior and the knowing disregard of illicit material distributed to and involving youths on Meta’s platforms. Additionally, Defendants’ acts

and practices tend to negatively affect competitive conditions by foreclosing or impairing the opportunities of market participants, limiting consumer choice and harming consumers.

477. Defendants’ acts and practices, including (i) Defendants’ repeated failure to act upon, inhibit, remove, or otherwise restrict access to illicit and/or illegal content constituting human trafficking and/or distribution or solicitation of CSAM, and (ii) Defendants’ failure to design the platforms to restrict such content, are offensive to public policy, as defined by statute and common law.

478. Moreover, Defendants’ design of their platforms to re-distribute and amplify CSAM and to facilitate connecting and monetizing networks of predators soliciting or distributing or seeking to distribute CSAM and/or engage in human trafficking are directly contrary to public policy that prohibit this trade.

479. The protection of minors and other New Mexico residents from the dangers of human trafficking and the associated mental and physical harm is a well-established objective underlying public policy. *See, e.g.*, NMSA 1978, § 30-52-1 (2008) (prohibiting human trafficking). The protection of minors and other New Mexico residents from the dangers of distribution or solicitation of CSAM and the associated mental and physical harm is a well-established objective underlying public policy. *See, e.g.*, NMSA 1978, §§ 30-6A-3(C) (2016) (prohibiting the distribution of CSAM), 30-37-3.2 (2007) (prohibiting “[c]hild solicitation by electronic communication device”); NMSA 1978, § 30-37-3.2 (2007) (prohibiting “soliciting a child under sixteen years of age, by means of an electronic communication device, to engage in sexual intercourse, sexual contact or in a sexual or obscene performance . . .”). Defendants’ acts and practices alleged herein—including Defendants’ failure to address illicit and illegal content and the users who distribute such content—therefore offend public policy.

480. Because Meta lacks and historically lacked effective age verification, Meta has obtained data from children under 13-years old in violation of public policy, because Meta failed to provide notice and seek consent from parents before it collected or used personal information from children. This constitutes an unfair practice under the UPA because the protection of children under the age of 13 from online abuse and the collection of their personal information is a well-established objective underlying public policy nationally and in New Mexico. To avoid any doubt, the State does not assert a claim pursuant to its authority to enforce the Children’s Online Privacy Protection Act (“COPPA”), but asserts instead that Meta’s practices in violation of COPPA constitute unfair practices under New Mexico law.

481. Specifically, and incorporating the allegations above, Zuckerberg’s role in making and approving decisions to: (i) design and maintain Facebook as addictive and to maximize the time users, including children and teenagers, spend on the platform; (ii) fail to address age limits and restrictions on content for the young users of Meta’s platforms; (iii) not to address, or even respond to, data regarding the bad experiences of teenager users on Meta’s platforms, including, but not limited to, sexual advances, suicide, bullying and harassment, depression, and body image issues; (iv) make dramatic cuts to the safety and moderation teams and failure to increase the size of those teams in response to executives’ requests, despite information that Meta already failed to identify, report, and address CSAM and trafficking on its platforms; (v) not take action to address Meta’s algorithm, despite knowledge that they amplified “egregious content” on the platform, such as CSAM, because of the “engagement cost” of such changes; (vi) rejecting a decision to hide like counts on posts in default settings and to maintain a ban on cosmetic surgery filters, and refusing to turn off notifications despite knowledge of the impact that these features had on teenage users’ mental health; (vii) deploying an ineffective “time spent” tool that fails to help teenagers limit their

use of the platforms; (viii) remove or maintain certain posts and content on Meta's platforms; and (ix) terminating research and surveys that turned up negative information and for directing Meta employees to not document critical findings and proposals to address these findings.

482. Defendants' unfair and/or unconscionable practices are the results of design features of its platforms, such as the operation of its algorithms in recommending users, groups and posts, the absence of effective age verification, the lack of separation between adults and minors, the failure to detect, remove, and report CSAM.

483. Defendants' unfair and/or unconscionable practices include, but extend beyond, developing an illegal market for inherently unlawful activity involved in obtaining and selling CSAM and the commercial sexual exploitation of children. This includes creating professional sites and allowing users engaged in illicit activities to sell advertising.¹⁵⁹ Defendants' acts and practices to induce young users' addictive and problematic use of their social media platforms are also immoral, unethical, oppressive, unscrupulous, and unconscionable. As described in detail in the foregoing paragraphs, Defendants, at all relevant times, based on their own internal research, had knowledge of the severe harms suffered by young users as a result of human trafficking, CSAM, the addictive use of their platforms and the role their platforms played in exacerbating those harms. Instead of taking meaningful measures to mitigate these damaging effects, Defendants knowingly, deliberately, and recklessly disregarded and turned a blind eye to them in pursuit of profit. Further, Defendants' willful design and use of platform tools and features to target, prey on, exploit, and manipulate highly vulnerable young users is unconscionable. Defendants' failure to warn of the dangers of their design choices and platform tools and features is unconscionable.

¹⁵⁹ <https://www.facebook.com/creators/tools/mta#check-eligibility>.

484. Defendants' acts and practices alleged herein also have caused and continue to cause unjustified substantial injury to consumers that could not be reasonably avoided. Namely, young users throughout New Mexico are suffering severe negative effects from addictive use of Defendants' platforms, including negative effects on sleep and school performance, emotional and behavioral challenges, poor mental health outcomes such as depression and anxiety, and negatively altered brain chemistry. Young users also could not have reasonably avoided the injuries resulting from Defendants' acts and practices, including because Defendants misrepresented and failed to disclose the dangerous nature of their social media platforms, and because Defendants utilized psychologically manipulative engagement-inducing features, knowing that young users are especially vulnerable to those psychologically manipulative tactics due to their lack of knowledge, ability, experience, or capacity.

485. The public health and safety risks and harm resulting from use Defendants' social media platforms are not outweighed by any countervailing benefit to consumers or competition.

486. But for these unfair and unconscionable practices, New Mexico consumers would not have incurred millions of dollars in damages, including without limitation the costs of treatment for mental and emotional trauma resulting from Defendants' actions and/or inaction, damages related to suicide and self-harm inflicted by youth and adolescents in New Mexico, and the societal costs attendant to human trafficking and solicitation/distribution of CSAM.

487. As a direct and proximate cause of the Defendants' unfair and/or unconscionable trade practices, New Mexico and New Mexico consumers have been injured in an amount to be determined at trial.

488. Defendants' unfair and/or unconscionable trade practices are willful and subject to a \$5,000 civil penalty for each and every violation per each Defendant. NMSA 1978, § 57-12-11 (1970).

489. Each unfair act by Defendants and/or each exposure of a New Mexico resident to illicit, illegal, or harmful content on Defendants' platforms resulting from the aforementioned conduct of each and all Defendants constitutes a separate violation of the Unfair Trade Practices Act.

490. New Mexico consumers and youth are suffering, have suffered, and will continue to suffer unjustified substantial injury as a result of Defendants' violations of New Mexico laws. Absent injunctive relief by this Court, Defendants are likely to continue to injure consumers and harm the public interest.

491. Plaintiff, State of New Mexico, seeks all legal and equitable relief as allowed by law, including inter alia injunctive relief, disgorgement of unjust profits, damages as allowed by law, all recoverable penalties under Section 57-12-11 including a civil penalty of \$5,000 per each violation per each Defendant named in this Count, attorney fees and costs, and pre- and post-judgment interest.

COUNT III
VIOLATION OF NEW MEXICO UNFAIR PRACTICES ACT
(UNCONSCIONABLE TRADE PRACTICES)
NMSA 1978, § 57-12-1 to -26
(Against All Defendants)

492. The State re-alleges all prior paragraphs of this Complaint as if fully set forth herein.

493. The Unfair Practices Act (“UPA”) prohibits “[u]nfair or deceptive trade practices and unconscionable trade practices in the conduct of any trade or commerce.” NMSA 1978 § 57-12-3 (1971).

494. Defendants are engaged in “trade” or “commerce” as defined by the UPA, which “includes the advertising, offering for sale or distribution of any services and any property and any other article, commodity or thing of value, including any trade or commerce directly or indirectly affecting the people of this state.” NMSA 1978, § 57-12-2(C) (2019). Defendants advertise, offer and distribute their internet platforms within New Mexico and to New Mexico residents.

495. In addition to offering, advertising, and distributing its social media platforms in New Mexico, Meta receives revenue both for showing ads to New Mexico consumers and also for harvesting New Mexican consumers’ personal data, including information about their activities and interests, to target advertising, thereby increasing its revenue from selling ads. Meta’s platforms also facilitate the sale of goods and services, both through advertisements that Meta directs to New Mexico residents and within New Mexico in exchange for a fee, and by providing space for users to offer, buy and sell merchandise (“Facebook Marketplace”). Meta charges and collects a fee when items are sold on its Marketplace, including in New Mexico. Meta enables users to monetize their accounts in order to sell subscriptions or permit advertisements to be placed on their platforms. Additionally, users may receive “stars” that other users purchase from Meta and those “stars” can be monetized by the user as well.

496. At all times relevant herein, the Defendants violated the UPA, NMSA 1978, §§ 57-12-1 to -26 (1967, as amended through 2019), by engaging in acts or practices in connection with the sale ... of any goods ... that to a person’s detriment, (1) takes advantage of the lack of knowledge, ability, experience or capacity of a person to a grossly unfair degree; or (2) results in

a gross disparity between the value received by a person and the price paid.” NMSA 1978 § 57-12-2E (2019).

497. The Attorney General is authorized to bring an action in the name of the State to remedy violations of the UPA. NMSA 1978, §§ 57-12-8(A) (1978), 57-12-15 (1967). This action is proper in this Court because Defendants are using, have used, and continue to use practices that are unlawful under the UPA. Section 57-12-8(A).

498. Defendants’ conduct, as alleged in the foregoing paragraphs, constitute unconscionable trade practices within the meaning of the Unfair Practices Act, including because Defendants made material statements, representations, omissions, and/or concealed information in a way that had the capacity or tendency to mislead consumers.

499. By engaging in the affirmative misrepresentations and omissions described above, Defendants took advantage of children’s and parents’ lack of knowledge, ability, experience or capacity in deciding when, whether, how, and how often to use Meta’s platforms. Without accurate information about the consequences to young users of using its platforms, New Mexico children, in particular, as well as their parents, could not make informed decisions about opening accounts on Instagram, Facebook, and WhatsApp, setting up account features, supervising or being supervised on the use of the platforms, and participating in groups or accepting friend requests, among other choices. The imbalance in information, experience, ability, and capacity between Meta, a multi-billion global corporation which extensively researched the activity on and effects of its platforms, and children using its platforms, was grossly unfair, and took advantage of their inferior knowledge of Meta’s products. Nor do children have the ability to assess Meta’s terms of services or features or to negotiate different terms of participation.

500. By agreeing to allow Meta to collect and use their data and to receive advertising, for which Meta was paid billions of dollars, Meta's users paid a price for access to its services. In addition, Meta sold goods and services in New Mexico not only by providing access to its platforms to millions of New Mexico consumers, but by selling paid advertising that was shown to New Mexico consumers who used its platforms and offering a marketplace for selling subscription content, listing jobs, or providing other goods and services (legal and illegal). By providing a product that subjected users, particularly young users, to the human trafficking, CSAM, solicitation, and other sexually explicit content, and to the compulsive use, depression, anxiety, eating disorders, negative self-worth, sleep disturbance, suicide, and other harms, the products that Meta delivered had a grossly disparate value.

501. Absent Meta's unfair, deceptive, and unconscionable conduct, many New Mexico consumers would not have used Meta's platforms and served as the targeted audience that allowed Meta to reap windfall profits.

502. Each unconscionable trade practice engaged in by Defendants as recited above and throughout this Complaint constitutes a separate violation of the Unfair Practices Act.

503. New Mexico consumers and youth are suffering, have suffered, and will continue to suffer unjustified substantial injury as a result of Defendants' violations of New Mexico laws. Absent injunctive relief by this Court, Defendants are likely to continue to injure consumers and harm the public interest.

504. Plaintiff, State of New Mexico, seeks all legal and equitable relief as allowed by law, including inter alia injunctive relief, disgorgement of profits, all recoverable penalties under Section 57-12-11 including a civil penalty of \$5,000 per each violation per each Defendant named in this Count, attorney fees and costs, and pre- and post-judgment interest.

COUNT IV
PUBLIC NUISANCE
NMSA 1978, § 30-8-81 and common law
(Against Defendants Meta Platforms, Inc., Instagram, LLC, Meta Payments, Inc. and
Meta Platforms Technologies, LLC)

505. The State re-alleges all prior paragraphs of this Complaint as if fully set forth herein.

506. The Attorney General may bring an action to abate a public nuisance in the name of the State. NMSA 1978, § 30-8-8(B) (1963).

507. Through the unreasonable and unlawful conduct described above, particularly in Counts 1 and II, Defendants, individually and in concert with each other, have contributed to, and/or assisted in creating and maintaining a condition that is harmful to the health and safety of thousands of New Mexico residents and interfered with the enjoyment of life in violation of New Mexico law.

508. In addition, Defendants' conduct contributing to the public nuisance was unreasonable in that it breached the duty Defendants assumed when they offered, marketed, and maintained their platforms without reasonable care and with defects that Defendants knew rendered them unsafe for children and assured and failed to warn children, their parents, and the public generally that their platforms were safe and that the CSAM, other CSEC and other harmful content on the platforms were extremely rare, and that children did not experience addiction or other mental health harms associated with their use of the platforms.

509. Internet-facilitated human trafficking, distribution of CSAM and other illicit material over the internet, and social media addiction and its impact on the social and mental well-being of New Mexico teens and adolescents are a public nuisance in New Mexico, which remains

unabated. The unlawful and unreasonable conduct by the Defendants has created and/or facilitated these hazards to public health and safety.

510. The health and safety of New Mexico’s children and others who use Meta’s platforms, as well as those impacted or affected by Meta’s platforms—i.e., teens or children suffering from the harmful effects of platform usage, is a matter of great public interest and of legitimate concern to the State’s citizens and residents.

511. The public nuisance created by Defendants’ actions is substantial and unreasonable – it has caused and continues to cause significant harm to the community, and the harm inflicted outweighs any offsetting benefit.

512. Defendants knew, or should have known, that the design and function of their internet platforms, including, but not limited to, the operation of their algorithms in promoting and encouraging illicit content related to human trafficking, CSAM, suicide, eating disorders, bullying or other topics known to cause harm to teens or adolescents, would create a public nuisance.

513. Defendants are liable for a public nuisance because they acted without lawful authority in knowingly creating and maintaining their platforms and their features, including, but not limited to, the algorithms that recommend content in such volumes to such a degree as to create an epidemic, which clearly affects a number of citizens, is injurious to public health, safety, morals and welfare, and interferes with the exercise and enjoyment of public rights. NMSA 1978, § 30-8-8-1.

514. Each Defendant is liable for public nuisance because their conduct at issue has caused an unreasonable interference with a right common to the general public. *City of Albuquerque v. State ex rel. Village of Los Ranchos de Albuquerque*, 1991-NMCA-015, ¶ 17, 111 N.M. 608 (“A public nuisance is a wrong that arises by virtue of an unreasonable interference with

a right common to the general public.”) (citing Restatement (Second) of Torts § 821B(1). The Defendants’ conduct described herein significantly interferes with public health, safety, peace, comfort, and convenience.

515. Defendants’ actions were, at the least, a substantial factor (i) enabling human trafficking to occur within New Mexico and affecting New Mexico residents, (ii) enabling the solicitation, distribution and creation of illicit sexual material involving children or child abuse, (iii) harming the well-being of numerous New Mexico teens and adolescents; (iv) causing addiction to social media; and/or (v) contributing to an increase in suicide, eating disorders, depression, bullying and other forms of harm among New Mexico teens and adolescents. Without Defendants’ actions, all of these harms resulting from use and abuse of Meta’s platforms would not have become so widespread, and the enormous public health hazard of social media addiction, including addiction to Meta’s platforms, human trafficking enabled by Meta’s platforms, distribution of child pornography enabled by Meta’s platforms, and increases in eating disorders, bullying and suicide among New Mexico teens enabled by Meta’s platform that now exists would have been averted.

516. In addition to the foregoing, Defendants’ conduct invades a legally protected interest. Defendants’ conduct constitutes an unreasonable interference because inter alia each Defendant has violated public policies intended to stem the tide of sexual exploitation of children and human trafficking. *See, e.g.*, NMSA § 30-52-1 (prohibiting human trafficking); NMSA § 30-37-3.2 (prohibiting “soliciting a child under sixteen years of age, by means of an electronic communication device, to engage in sexual intercourse, sexual contact or in a sexual or obscene performance . . .”).

517. Because Defendants have maintained their social media platforms contrary to law, and because Defendants' conduct has unreasonably interfered with a right common to the general public, Defendants are liable for public nuisance per se. See *Espinosa v. Roswell Tower, Inc.*, 1996-NMCA-006, ¶ 10, 121 N.M. 306, 910 P.2d 940 ("An activity conducted or maintained contrary to law may be a public nuisance per se when the activity unreasonably interferes with a right common to the general public.").

518. Defendants' unreasonable interference with a right common to the public is of a continuing nature.

519. Defendants are aware of the unreasonable interference that their conduct has caused in the State of New Mexico. Internal documents described above demonstrate the Defendants' knowledge of the harms their conduct was causing to society at large, including to teens and adolescents in New Mexico. Defendants were aware of and actively monitored scores of news reports providing evidence of their users suffering harm as the result of the design of their platforms.

520. The public nuisance created by Defendants' actions is substantial and unreasonable – it has caused and continues to cause significant harm to the community, and the harm inflicted outweighs any offsetting benefit. Incidents of human trafficking, distribution or solicitation of CSAM and human trafficking, and youth suicide, eating disorders, bullying, and depression are widespread throughout New Mexico and have caused harm to the entire community that includes, but is not limited to:

- a. Increase in the rate of suicides, depression, eating disorders, and other mental health issues among young people in New Mexico attributable to social media addiction and misuse;

- b. Increased incidences of human trafficking occurring in New Mexico or affecting New Mexico residents facilitated by Defendants' platforms;
- c. Increase in the decline of physical and mental well-being among young people in New Mexico attributable to social media addiction and misuse, and the attendant societal and economic costs associated therewith;
- d. Increase in creation and distribution of, and exposure to CSAM by teens and adolescents, and the attendant societal and economic costs associated therewith;
- e. Decline in educational attainment by teens and adolescents due to loss of sleep or other effects from overuse or misuse of Defendants' platforms;
- f. Long-term effects from eating disorders and body dysmorphia, including dermatological effects to the nails and hair, gastrointestinal illnesses, fertility issues, and impacts to the endocrine system, nervous system and skeletal system; and
- g. Increase in harms resulting from the overuse and abuse of Defendants' platforms, including dissociative behavior, withdrawal symptoms, social isolation, damage to body image and self-worth, increased risky behavior, exposure to predators, sexual exploitation, and other profound mental health issues.

521. Plaintiff, the State of New Mexico, seeks all legal and equitable relief as allowed by law, including inter alia injunctive relief, abatement of the public nuisance, payment to the State of monies to abate the public nuisance, attorney fees and costs, and pre- and post-judgment interest.

XX. PRAYER FOR RELIEF

WHEREFORE, the State of New Mexico, by and through its Attorney General, respectfully prays that this Court grant the following relief:

1. Entering judgment in favor of the State in a final order against Defendants;

2. Declaring that each act, statement and/or omission of Defendants described in this Complaint constitute separate and willful violations of the UPA;
3. Declaring that Defendants' unreasonable and unlawful conduct created a public nuisance;
4. Imposing civil penalties on each Defendant of up to \$5,000 for each violation of the UPA;
5. Permanently enjoining Meta and its employees, officers, directors, agents, assigns, successors, subsidiaries, and other persons acting in concert or participation with it, from engaging in unfair, unconscionable, or deceptive practices in violation of New Mexico law and ordering a permanent injunction;
6. An order that Meta abate the public nuisance caused by Meta's unreasonable and/or unlawful conduct;
7. Disgorgement of profits and data that were unjustly obtained;
8. The cost of investigation, reasonable attorneys' fees, and all costs and expenses;
9. Pre-judgment and post-judgment interest; and
10. All other relief as provided by law and/or as the Court deems appropriate and just.

Plaintiff asserts claims herein in excess of the minimum jurisdictional requirements of this Court.

RESPECTFULLY SUBMITTED this 5th day of December 2023.

RAÚL TORREZ
ATTORNEY GENERAL OF NEW MEXICO

/s/ James W. Grayson
JAMES W. GRAYSON
CHIEF DEPUTY ATTORNEY GENERAL
New Mexico Attorney General's Office
P.O. Drawer 1508
Santa Fe, NM 87504-1508
Phone: (505) 218-0850
Email: jgrayson@nmag.gov

SERENA R. WHEATON
ASSISTANT ATTORNEY GENERAL

New Mexico Attorney General's Office
Consumer & Environmental Protection Division
201 Third St. N.W., Suite 300
Albuquerque, NM 87102
Phone: (505) 490-4846
Email: swheaton@nmag.gov

LINDA SINGER
DAVID I. ACKERMAN
MOTLEY RICE LLC
401 9th St., NW, Suite 630
Washington, D.C. 20004
Phone: (202) 232-5504
Email: lsinger@motleyrice.com
Email: dackerman@motleyrice.com
(*Pro Hac Vice Petitions Forthcoming*)

EXHIBIT 1



Bad Experiences and Encounters Framework (BEEF) Survey

Data collection: June 27-July 8, 2021

[REDACTED] (UXR)

Signals and Insights Platform



Background

Why are we doing this?

To develop a holistic, consistent picture of user bad experiences on Instagram that allows us to track our progress each half.

Holistic: TRIPS focuses on policy-violating experiences, but there are additional causes of bad experiences on Instagram. By contrast, the BEEF Survey pulls issues from the Bad Experiences and Encounters Framework, which sourced issues from all of our user input channels.

Consistent: We have many sources of valuable user data, such as individual UXR surveys and interviews, FRX reports, flytrap, app store reviews, and social media listening. But, it's hard to make apples-to-apples comparisons between these data points, because of differences in how they're collected and weighted. The BEEF Survey asks a random sample of users the top 22 issues from BEEF, so we can compare and contrast easily.

Track our progress: Without a control group, it's impossible to determine causality from our other signals. The BEEF Survey was fielded to both the well-being holdout and production groups.

Survey timeline

Date	To do	Status
April 12	Agree on which TRIPS questions to include	Complete
April 19	Use framework project results to generate proposal for additional survey constructs	Complete
April 26	Develop survey questions/ look for questions from existing surveys	Complete
May 3	Develop survey questions/ look for questions from existing surveys	Complete
May 10	Submit to research review	Complete (June 3)
May 24 June 3	Submit to survey review	Complete (June 9)
May 31 June 10	Submit to translations	Complete (June 29)
June 14 June 30	Launch survey	Complete (June 30)
June 21 July 6	Data cleaning	Complete
June 28 August 25	Data weighting/ appending	Complete (August 25)
July 19 August 26	Data cleaning	Complete (September 6)
July 26 September 7	Analysis/ reporting	In progress
October 4	Socialization	
October 11	v2 Framework	
October 18	Begin BEEF Survey v2 development	
December 16	Launch BEEF Survey v2 (code freeze Dec 16-Jan 4)	

What’s next?

The plan is to add findings to this deck and the Bad Experiences Working Group as I go, with the goal of having a draft completed in the next two weeks. After that, I will focus on 1:1 and small group meetings with various teams, to share the findings relevant to them.

As I engage in socialization of the findings, I will also begin a literature review of work done since the v1 BEEF, to see if any new issues have emerged that should be included in the next BEEF Survey (v2 Framework, although it will be much more lightweight than the v1 effort).

I will use the current BEEF Survey and any new insights gathered during the v2 Framework to begin the survey development process for the v2 BEEF Survey.

Survey questions, issues

Questions that are also in TRIPS are highlighted in yellow. All questions had the same response options as TRIPS: ‘Yes, during the last 7 days,’ ‘Yes, but more than 7 days ago,’ and ‘No.’

False or misleading	Have you ever seen anything on Instagram that was false or misleading?
Violence	Have you ever seen any violent, bloody, or disturbing images on Instagram that bothered you?
Hate (witness)	Have you ever seen anyone discriminating against people on Instagram because of their gender, religion, race, sexual orientation, or another part of their identity?
Bullying (witness)	Have you ever seen anyone do any of these things to someone else on Instagram? -Insult or disrespect them -Contact them in an inappropriate way -Damage their reputation -Threaten them -Exclude them or leave them out
SRG drug sales	Have you ever seen anyone trying to buy or sell any of the following things on Instagram? -Prescription drugs -Illegal drugs
Bullying (target)	Has anyone ever done any of these things to you on Instagram? -Insulted or disrespected you -Contacted you in an inappropriate way -Damaged your reputation -Threatened you -Excluded you or left you out
Nudity	Have you ever seen nudity or sexual images on Instagram that you didn't want to see?
Negative Social Comparison	Have you ever felt worse about yourself because of other peoples’ posts on Instagram?

Unwanted advances	Have you ever received unwanted sexual advances on Instagram?
Account security	Have you ever been unable to access your Instagram account, either because of a lost password or because your account was hacked?
Civic content	Have you ever felt that you have seen too many political posts on Instagram?
Impersonation	Have you ever found out that a Instagram account was pretending to be you?
Self-harm	Have you ever seen someone harm themselves, or threaten to do so, on Instagram?
Spam/ fake account contact	Have you ever been contacted on Instagram by an account that seemed fake?
Over enforcement	Have you ever felt that Instagram was wrong when it removed a post or video for violating the rules?
Audience limitation	Have you felt that not enough people see the things you share on Instagram?
Usability, action	Have you ever had difficulty posting a picture or video to Instagram?
Transparency	Have you ever felt Instagram isn't transparent about why it removes posts that are against the rules?
Usability, consumption	Have you ever experienced the Instagram app freezing, crashing, or closing unexpectedly?
Data privacy	Have you ever had concerns about the way Instagram might use data and information about you?
Perceived control	Have you ever felt a lack of control over what you see on Instagram?
Commerciality	Have you ever felt that you have seen too many ads on Instagram?

Survey questions and log data, context

All questions with response options can be found [here](#).

Question name	Question text
Surface	In the last 7 days, where in Instagram did you see this? Please select all that apply.
Frequency	In the last 7 days, how many times have you seen something like this?
Specific emotion felt	Which specific emotions did you feel? Please select all that apply.
Length of emotional reaction	For how long after the experience did you feel this way?
Perceived support	Thinking about this experience, did you feel supported by Instagram?
Action taken (both on and off app)	Did you take any of the following actions as a result of this experience? Select all that apply.
Stops posting?	Did this experience stop or discourage you from posting on Instagram?
Known person, online/offline	What's your relationship with the account that posted this?
Age (13-15, 16-17, 18...)	What is your age?

Log data	Name	Table
Country	country	dim_ig_users:instagram
App use	interface_l28: instagram_classic	dim_ig_users:instagram
Gender	gender	dim_ig_users:instagram
Followers	followers	dim_ig_users:instagram
Public or private account	is_private	dim_ig_users:instagram
Feed posts	feed_media_7d	dim_ig_creators:instagram
Reels posted	reels_media_7d	dim_ig_creators:instagram
Stories posted	story_media_7d	dim_ig_creators:instagram
Public media posted	public_media_7d	dim_ig_creators:instagram
Creator segmentation	ig_creator_segment	dim_ig_creators:instagram
iOS vs Android	primary_interface	dim_ig_user_primary_device:instagram
RAM	ram_class	dim_ig_user_primary_device:instagram
Hi NAC views*	viewer_high_nac_vpvs_l7	ig_user_social_comparison_metrics:instagram
Overall views	viewer_vpvs_w_prediction_l7	ig_user_social_comparison_metrics:instagram
Sessions^	num_sessions	ig_user_app_session_counts:instagram
time spent^	time_spent	ig_user_app_session_counts:instagram
meaningful impressions^	sum_meaningful_engagement_viewer_score	ig_meaningful_engagement_user_agg:instagram

Sample size by issue

The sample size was calculated to detect a 1% absolute difference between the two groups (production and holdout). Part of that calculation is the baseline prevalence of each issue; past research was used to provide those estimates. If an issue wasn’t actively being worked on by a team in the IG Community Pillar in H1 2021, then that issue was only asked in the production group.

Each respondent was asked about five issues (randomly chosen from the set of 22). If they reported experiencing at least one of the five issues, the survey system randomly chose one of the issues they said yes to and asked a series of follow-up questions. If they said they didn’t experience any of the five issues, they were asked their age followed by a series of questions unrelated to this report.

Issue	Baseline Prevalence	Detectable drop	n per group	# of groups	Subtotal, planned	Subtotal, actual
Account security (access)	20.00%	19.00%	24,641	1	24,641	26,907
Audience limitation	33.70%	32.70%	34,813	2	69,626	71,845
Bullying and harassment, target	7.80%	6.80%	10,622	2	21,244	24,790
Bullying and harassment, witness	22.43%	21.43%	26,875	2	53,750	57,744
Commerciality	21.50%	20.50%	26,042	1	26,042	28,498
Data privacy	31.90%	30.90%	33,813	2	67,626	69,802
Drugs and Related Goods	3.64%	2.64%	4,774	2	9,548	12,242
False or Misleading	26.68%	25.68%	30,337	2	60,674	64,854
Graphic Violence	11.30%	10.30%	15,122	2	30,244	34,928
Hate Witness	19.73%	18.73%	24,381	2	48,762	53,283
Impersonation (first person)	3.10%	2.10%	301	1	301	5,880
Negative Social Comparison	20.00%	19.00%	24,641	2	49,282	53,954
Nudity	13.46%	12.46%	17,707	2	35,414	40,571
Over enforcement	30.00%	29.00%	32,647	2	65,294	68,226
Perceived control / sense of place	23.10%	22.10%	27,458	1	27,458	29,774
Political posts	28.20%	27.20%	31,437	1	31,437	33,516
Unwanted advances / solicitation	30.00%	29.00%	32,647	2	65,294	69,146
Self-harm (witness)	10.00%	9.00%	13,495	1	13,495	15,830
Spam (fake account)	50.00%	49.00%	39,240	1	39,240	40,629
Transparency	30.00%	29.00%	32,647	2	65,294	66,966
Usability / action-oriented	21.30%	20.30%	25,859	2	51,718	57,160
Usability / consumption-oriented	26.20%	25.20%	29,974	2	59,948	67,836
			question <i>n</i> =		916,332	994,381
			respondent <i>n</i> =		238,246	237,923

Planned analyses (green = drafts complete)

1. Issue prevalence for holdout vs prod, and overall

2. Issue prevalence by subgroups

- a. Age
- b. Creators
- c. Gender

3. Add frequency, surface, and other context data to the issue prevalence tables

4. Additional cuts by subgroups

- a. RAM class/interface
- b. NAC (negative experience content) I7
- c. App use (number of sessions, time spent)
- d. Meaningful impressions
- e. Surface

5. Issue-specific analyses. 22 issues by 9 follow-up questions—I'll start with issues that our teams are actively working on, then go by prevalence

6. Grouping of issues. Analyze the ways individual issues cluster (or don't) with each other. Assuming I find one or more commonalities between issues, I'll use those to create rubrics to understand the full slate of issues users experience



Results

Issue rates, full dataset

Improvements in perceived violence and usability action, but a decline in perceived data privacy

The issues in the table to the right are ranked by the percentage of respondents in the full dataset who selected “Yes, during the last 7 days.” Other response options were “Yes, but more than 7 days ago” and “No.” Those responses are not displayed here for space reasons. Production refers to respondents in the production version of IG, while Holdout refers to respondents in the Community Pillar holdout group in H1. If an IG team was not actively working on an issue in H1, the survey question was only asked to the production condition.

Statistical tests for differences in column proportions between the production and holdout groups were conducted, with and without a Bonferroni correction for alpha error inflation (no difference in results).

There was a statistically significant improvement in perceptions of violence (“Have you ever seen any violent, bloody, or disturbing images on Instagram that bothered you?” and usability action (“Have you ever had difficulty posting a picture or video to Instagram?”) between production and holdout.

There was also a statistically significant decline in perceptions of data privacy (“Have you ever had concerns about the way Instagram might use data and information about you?” between production and holdout.

	Overall		Production		Holdout	
	Column N %	n	Column N %	n	Column N %	n
commerciality	48.2%	11,493	48.2%a	11,493	--	
misinfo	30.3%	19,879	30.6%a	8,253	30.1%a	11,626
bully witness	28.3%	16,567	28.1%a	6,685	28.5%a	9,882
audience limitation	26.8%	19,545	26.8%a	8,058	26.8%a	11,488
usability passive	25.5%	17,344	25.6%a	7,626	25.5%a	9,718
hate witness	25.3%	13,710	25.1%a	5,511	25.5%a	8,199
data privacy	24.4%	17,293	24.8%b	7,277	24.1%a	10,016
perceived control	23.9%	5,904	23.9%a	5,904	--	
fake acct 1st	21.8%	7,417	21.8%a	7,417	--	
negative comparison	19.2%	10,549	19.3%a	4,311	19.2%a	6,237
transparency	17.5%	11,931	17.5%a	4,896	17.5%a	7,036
political posts	17.0%	4,779	17.0%a	4,779	--	
nudity	16.3%	6,751	16.2%a	2,688	16.4%a	4,063
over enforcement	14.8%	10,206	14.8%a	4,239	14.7%a	5,967
violence	12.8%	4,580	12.3%b	1,764	13.2%a	2,816
unwanted advances	11.9%	8,364	12.0%a	3,486	11.8%a	4,878
usability action	10.7%	6,208	10.4%b	2,481	10.9%a	3,727
bully target	8.1%	2,055	7.8%a	802	8.2%a	1,254
self harm	6.7%	894	6.7%a	894	--	
acct security	3.9%	880	3.9%a	880	--	
drugs	3.9%	494	4.0%a	201	3.9%a	292
impersonation 1st	3.7%	178	3.7%a	178	--	

Note: Values in the same row and subtable not sharing the same subscript are significantly different at p< 0.05 in the two-sided test of equality for column proportions. Cells with no subscript are not included in the test. Tests assume equal variances. All significant findings remain significant when using a Bonferroni correction for alpha error inflation.

Issue rates and frequencies in last 7 days

Commerciality and political posts have the highest frequencies among respondents who have experienced the issue, while fake accounts and impersonation have the lowest.

If a respondent experienced an issue (i.e., responded “Yes, during the last 7 days”), they were then asked *how often* they experienced that issue. The percentage of respondents selecting each of the answer options is displayed in the table, along with the average frequency (first yellow column).. The issues (rows) are ranked by average frequency, with the most frequent issues on top. Commerciality and political posts have the highest frequencies, while impersonation and fake account contact have the lowest.

The second yellow column is a rate x frequency composite score (the rate of an issue being experienced in the dataset, multiplied by the average frequency). Commerciality also has the highest score on this metric, followed by bullying witness; impersonation is the lowest.

Finally, the last two columns rank the issues by issue rate (penultimate column) and issue x frequency composite (last column). Highlighted in blue are the issues with the biggest deltas: the largest is political posts, which ranks 12th by issue rate, but 7th by issue rate x frequency.

Issue	Rate	Frequency, last 7 days							Rank	
		Once	2-3	4-5	6-9	10+	Average freq*	Rate x Freq	Issue	Issue x freq
commerciality	48.2%	5.3%	22.1%	28.8%	15.0%	28.8%	5.90	2.85	1	1
political posts	17.0%	7.3%	31.8%	27.4%	10.0%	23.4%	5.20	0.88	12	7
nudity	16.3%	18.0%	39.3%	17.0%	6.8%	19.0%	4.33	0.71	13	11
perceived control	23.9%	15.7%	38.4%	22.2%	7.2%	16.5%	4.31	1.03	8	6
hate witness	25.3%	16.9%	41.4%	18.2%	5.8%	17.6%	4.22	1.07	6	5
negative comparison	19.2%	17.9%	39.1%	19.3%	7.2%	16.6%	4.22	0.81	10	9
audience limitation	26.8%	25.4%	34.9%	17.5%	6.0%	16.2%	3.99	1.07	4	4
bully witness	28.3%	23.0%	39.6%	16.0%	4.8%	16.6%	3.96	1.12	3	2
misinfo	30.3%	21.7%	46.3%	15.3%	4.4%	12.3%	3.62	1.10	2	3
usability passive	25.5%	21.5%	46.0%	18.5%	5.6%	8.4%	3.46	0.88	5	8
violence	12.8%	27.4%	42.9%	14.6%	4.5%	10.6%	3.40	0.44	15	15
drugs	3.9%	34.1%	37.6%	11.3%	5.1%	12.0%	3.36	0.13	21	20
self harm	6.7%	33.9%	38.3%	12.3%	4.3%	11.1%	3.28	0.22	19	19
bully target	8.1%	38.3%	34.0%	12.0%	3.9%	11.8%	3.24	0.26	18	18
transparency	17.5%	35.7%	37.9%	11.7%	3.5%	11.2%	3.21	0.56	11	13
unwanted advances	11.9%	33.5%	40.1%	12.9%	5.0%	8.4%	3.14	0.37	16	16
acct security	3.9%	50.4%	22.9%	10.1%	2.3%	14.3%	3.13	0.12	20	21
data privacy	24.4%	38.4%	35.2%	12.3%	3.9%	10.3%	3.13	0.76	7	10
over enforcement	14.8%	36.1%	37.9%	12.7%	3.6%	9.8%	3.13	0.46	14	14
usability action	10.7%	34.0%	40.1%	13.5%	3.5%	9.0%	3.11	0.33	17	17
fake acct 1st	21.8%	38.9%	40.9%	10.8%	3.4%	6.1%	2.76	0.60	9	12
impersonation 1st	3.7%	54.1%	25.5%	7.0%	3.8%	9.6%	2.75	0.10	22	22

* In order to calculate the average, the frequency range selected by the respondent was converted into a single number: Once = 1, 2-3 = 2.5, 4-5 = 4.5, 6-9 = 7.5, and 10+ = 10.

Issue rates by surface

DM/chat has the two highest rates of issues across surfaces: 73.1% of the time for fake account contact, and 68.6% of the time for unwanted sexual advances.

If a respondent experienced an issue, they were asked where in Instagram it happened. The question allowed for multiple responses, so the rows will add up to more than 100%. Higher numbers are shaded in darker reds, and the highest number per issue/row is in bold.

A few patterns are clear--first, some surfaces are more likely to have issues occur than others, but no app surfaces are immune to bad experiences. Surface comparisons, however, should be taken with a grain of salt--people spend more time on some surfaces than others, so we'd expect differences just because of that.

Second, specific issues happen more often on specific surfaces. When respondents experienced fake account contact in the last 7 days, they said it occurred via DM/chat 73.1% of the time; the next highest is someone's profile, at 14.6%. By contrast, when respondents reported seeing self harm in the last 7 days, it was spread across various surfaces: witnessed in post comments 12.6% of the time, and Feed/Stories 31.9% of the time.

In the last 7 days, where in Instagram did you see this? Please select all that apply.

	DM/ Chat	Feed/ Stories	Comment on my post	Explore/ Search	Hashtag/ Location Page	Reels	Someone's Profile	Somewhere else	I don't remember
fake accounts	73.1%	7.6%	10.5%	4.9%	4.2%	3.6%	14.6%	10.3%	16.0%
unwanted advances	68.6%	9.2%	7.4%	9.2%	5.5%	6.0%	16.7%	10.6%	12.9%
bully target	60.1%	17.6%	18.5%	12.1%	12.0%	10.2%	20.5%	15.2%	17.7%
impersonation 1st	26.0%	23.0%	21.9%	24.0%	21.3%	19.5%	35.4%	27.2%	35.8%
nudity	23.4%	26.4%	7.9%	37.6%	18.1%	22.7%	37.0%	12.8%	18.6%
usability action	20.8%	46.6%	11.1%	14.5%	8.9%	17.0%	16.4%	13.6%	30.6%
usability passive	20.7%	49.7%	7.5%	21.4%	7.6%	19.9%	15.6%	12.2%	29.1%
acct security	18.4%	24.0%	11.2%	15.1%	12.5%	12.6%	18.0%	16.2%	55.2%
misinfo	18.2%	37.3%	6.4%	38.3%	11.5%	17.4%	24.1%	13.3%	21.2%
bully witness	17.8%	30.6%	10.1%	24.4%	13.3%	16.7%	40.8%	17.2%	23.2%
self harm	14.9%	31.9%	12.6%	25.3%	14.4%	19.1%	29.4%	13.8%	26.9%
drugs	14.8%	34.8%	11.7%	22.5%	17.6%	16.4%	26.6%	13.5%	28.3%
perceived control	13.7%	37.9%	5.8%	37.8%	12.2%	25.9%	13.0%	9.6%	25.5%
data privacy	13.5%	26.3%	5.8%	21.7%	10.3%	11.9%	11.9%	14.3%	46.8%
violence	11.3%	34.0%	7.2%	34.5%	16.2%	20.2%	29.7%	11.2%	23.1%
transparency	10.1%	28.4%	7.5%	14.5%	10.4%	11.2%	35.6%	11.5%	34.5%
hate witness	9.5%	31.0%	9.9%	31.3%	17.5%	20.2%	39.4%	19.6%	21.9%
audience limitation	9.3%	44.4%	9.2%	13.0%	10.9%	9.9%	13.0%	7.3%	41.3%
negative comparison	8.7%	40.9%	6.4%	36.3%	12.3%	24.7%	34.8%	9.0%	22.4%
over enforcement	7.0%	35.7%	6.5%	12.1%	9.6%	11.6%	42.7%	11.2%	29.7%
commerciality	6.9%	64.7%	5.0%	20.6%	10.2%	17.1%	8.9%	7.4%	23.4%
political posts	6.2%	44.4%	6.6%	34.0%	16.3%	20.0%	23.7%	10.4%	27.8%

An expanded table with 95% confidence intervals can be found [here](#).

Relationship between respondent and antagonist

Issues are largely caused by strangers, with negative social comparison (15.5%) and bullying target (14.6%) having the highest rates of known instigators.

If respondents experienced an issue, they were then asked, “*What’s your relationship with the account that posted this?*” (This follow-up question wasn’t asked for issues that wouldn’t make sense, like commerciality or over-enforcement.)

The content labeled an issue by the respondent was usually posted by someone they don’t know. 93.8% of unwanted sexual advances are from people the respondent doesn’t know, the highest percentage of the 12 issues; negative social comparison has the lowest, at 64.9%.

Conversely, the highest rates of issues instigated by someone the respondent knows offline were negative comparison (15.5%) and bullying target (14.6%). No other issue is above 10%. The lowest is unwanted sexual advances, with only 2.1% from people the respondent knows offline.

	I know them offline/ in real life	I know them, but only online	I don't know them
Negative comparison	15.5%	19.6%	64.9%
Bullying target	14.6%	16.0%	69.4%
Impersonation	9.9%	21.1%	69.0%
Drugs	9.7%	12.3%	78.0%
Self harm	9.5%	19.2%	71.3%
Political posts	8.9%	16.2%	75.0%
Bullying witness	6.4%	17.6%	76.0%
Violence	5.1%	13.7%	81.2%
Hate witness	4.6%	14.3%	81.1%
Misinfo	4.5%	9.7%	85.8%
Nudity	3.7%	9.1%	87.1%
Unwanted sexual advances	2.1%	4.2%	93.8%

An expanded table with 95% confidence intervals can be found [here](#).

Relationship with antagonist by creator status

Low response rates among creators prevent strong conclusions

In the table to the right, responses to the question “*What's your relationship with the account that posted this?*” are split by creator status (the “*I don’t know them*” response option is excluded for the sake of space). The rates of knowing the antagonist offline is highlighted in blue.

Given the very low numbers of creators (especially aspiring+ creators) who answered the known antagonist question, it’s not possible to make confident statements about the relationship between creator status and knowing the antagonist.

I also looked at the relationship between knowing the antagonist* and number of followers/number of stories posted in the last seven days (not official proxies for creator status, but they do correlate with creator status--see the creator segmentation descriptives slide). There aren’t any statistically significant correlations between follower count and knowing the antagonist. There are, however, multiple significant correlations between L7 stories production and knowing the antagonist for bullying witness ($r = 0.048$), hate witness ($r = 0.025$), misinfo ($r = 0.042$), self harm ($r = 0.12$), and negative social comparison ($r = 0.046$).

	I know them...	not creator		experimenting		aspiring+	
		%	n	%	n	%	n
Bullying target	only online	16.2%	227	15.0%	19	3.4%	1
	offline/ in real life	15.0%	210	11.9%	15	5.7%	1
Bullying witness	only online	17.2%	1479	23.8%	145	20.5%	8
	offline/ in real life	6.4%	554	6.0%	37	2.5%	1
Political posts	only online	15.7%	250	21.4%	23	38.7%	2
	offline/ in real life	9.0%	142	7.3%	8	13.7%	1
Drugs	only online	12.0%	38	17.0%	4	0.0%	0
	offline/ in real life	9.5%	30	12.5%	3	0.0%	0
Hate witness	only online	14.0%	1227	19.0%	115	12.1%	5
	offline/ in real life	4.6%	400	6.1%	37	2.6%	1
Violence	only online	13.5%	431	16.6%	36	5.2%	1
	offline/ in real life	5.1%	163	5.7%	12	0.0%	0
Unwanted sexual advances	only online	4.1%	84	5.0%	6	0.0%	0
	offline/ in real life	2.1%	42	2.1%	3	0.0%	0
Negative social comparision	only online	19.1%	908	27.8%	85	22.2%	4
	offline/ in real life	15.3%	727	17.6%	54	20.8%	4
Self harm	only online	19.3%	107	17.7%	8	23.7%	1
	offline/ in real life	9.5%	53	9.1%	4	15.0%	1
Nudity	only online	8.8%	404	14.6%	42	13.7%	3
	offline/ in real life	3.7%	171	4.3%	12	9.5%	2
Misinfo	only online	9.7%	786	10.4%	48	8.2%	3
	offline/ in real life	4.5%	367	4.9%	23	4.5%	1
Impersonation 1st	only online	21.4%	26	15.8%	1	0.0%	0
	offline/ in real life	9.9%	12	13.2%	0	0.0%	0

An expanded table with confidence intervals can be found [here](#).

*for the purposes of the correlation analysis, the known antagonist variable was coded: 1 = I don’t know them, 2 = only online, and 3 = offline.

Actions taken as a result of experiencing an issue

Blocking or muting the account is the most frequent action taken as a result of an issue experience, with closing Instagram as the second most frequent.

If a respondent reported experiencing an issue in the last 7 days, they were asked if they took any actions as a result. The table to the right shows the percentage of people who reported taking each action, with larger percentages shaded in darker reds. Actions (columns) are arranged from least to most frequently used.

Changing privacy settings is the least-used action across issues, with one outlier, first-person impersonation (29%). Talked to a parent or close friend was the second least-used action. Blocking or muting the account was the most frequent action, followed by closing Instagram.

Among issues, impersonation elicited the most actions overall, followed by bullying target and bullying witness.

Did you take any of the following actions as a result of this experience? Select all that apply.

	Changed my privacy settings	Talked to a parent or close friend about it	Deleted the post or comment	Reported the post, Story, or Reel	Unfollowed the account	Reported the account	Closed the Instagram app	Blocked or muted the account
Bullying target	12.0%	12.0%	22.9%	12.9%	28.5%	37.6%	25.7%	62.9%
Bullying witness	8.1%	12.5%	10.3%	18.3%	32.7%	39.3%	36.9%	37.6%
Drugs	8.2%	9.8%	12.0%	13.2%	30.4%	27.3%	32.9%	36.0%
Fake account	7.4%	3.8%	20.3%	6.2%	17.7%	28.7%	14.5%	68.7%
Hate witness	5.3%	14.4%	8.3%	21.0%	31.5%	38.1%	39.4%	32.1%
Impersonation 1st	29.0%	17.8%	18.8%	21.5%	23.7%	44.9%	24.0%	39.6%
Misinfo	5.8%	8.7%	9.5%	16.3%	32.4%	30.6%	35.5%	35.5%
Negative social comparison	7.4%	12.0%	6.7%	8.8%	29.2%	15.0%	54.8%	24.4%
Nudity	8.6%	7.7%	11.9%	20.8%	29.3%	36.1%	35.6%	42.0%
Political posts	6.6%	12.0%	9.1%	7.9%	31.0%	11.9%	45.9%	26.7%
Self harm	10.4%	14.9%	10.6%	16.2%	28.0%	29.8%	41.9%	32.3%
Unwanted sexual advances	8.6%	4.8%	21.8%	7.6%	19.3%	30.9%	17.6%	66.5%
Violence	6.3%	13.0%	9.7%	22.0%	27.1%	34.0%	36.6%	31.3%

An expanded table with confidence intervals can be found [here](#).

Perceived support from Instagram

Respondents felt the least supported by Instagram when they were the target of bullying, had difficulty posting, or felt Instagram took down content when it shouldn't.

If a respondent reported experiencing an issue in the last 7 days, they were asked if they took any actions as a result. The table to the right shows the percentage of people who reported taking each action, with larger percentages shaded in darker reds. Actions (columns) are arranged from least to most frequently used.

Changing privacy settings is the least-used action across issues, with one outlier, first-person impersonation (29%). Talked to a parent or close friend was the second least-used action. Blocking or muting the account was the most frequent action, followed by closing Instagram.

Among issues, impersonation elicited the most actions overall, followed by bullying target and bullying witness.

Thinking about this experience, did you feel supported by Instagram?				
	Completely/ mostly supported	Somewhat Supported	Not very/ not at all supported	Does not apply to my situation
Bullying target	16.5%	12.3%	49.2%	22.0%
Usability action	15.5%	13.1%	45.9%	25.6%
Over enforcement	12.5%	14.2%	42.9%	30.4%
Impersonation	27.6%	14.6%	41.4%	16.5%
Perceived control	12.3%	14.6%	41.1%	32.0%
Account security	25.6%	16.3%	40.0%	18.0%
Hate witness	13.8%	15.6%	39.5%	31.2%
Nudity	18.6%	16.2%	38.9%	26.3%
Violence	19.8%	14.0%	38.9%	27.3%
Self harm	19.5%	12.4%	38.0%	30.1%
Audience limitation	17.4%	15.8%	37.9%	28.8%
Transparency	16.6%	14.7%	37.6%	31.0%
Drugs	21.5%	13.1%	37.0%	28.4%
Bullying witness	14.7%	15.0%	36.9%	33.4%
Usability passive	16.0%	14.8%	36.6%	32.6%
Data privacy	17.2%	17.0%	35.0%	30.8%
Misinfo	15.1%	14.1%	34.5%	36.4%
Negative social comparison	14.7%	13.7%	34.1%	37.6%
Unwanted sexual advances	23.0%	15.4%	33.3%	28.3%
Commerciality	15.2%	14.2%	32.9%	37.7%
Political posts	18.2%	15.1%	29.4%	37.4%
Fake account	20.8%	15.5%	27.2%	36.5%

An expanded table with confidence intervals can be found [here](#).

Impact of issue experience on creation behavior

All issues cause some reduction in self-reported creation behavior, with impersonation hindering creation the most.

If a respondent experienced an issue, they were asked if it stopped or discouraged them from posting on Instagram. The table to the right shows the percentage of respondents selecting each of response options; issues are sorted by the percentage of respondents who selected 'yes' or 'somewhat.'

First-person impersonation, usability action, and account security are the top three, which makes sense; they all involve being physically unable to access or use your account. Negative comparison is #4, followed by over enforcement and audience limitation.

Contact by a fake account and unwanted advances are the issues that reduce creation the least (it's worth noting, however, that even the issue with the least impact on creation, fake account contact, resulted in 22.6% of respondents choosing 'yes' or 'somewhat.')

Did this experience stop or discourage you from posting on Instagram?

	Yes	Somewhat	No
impersonation 1st	35.9%	31.0%	33.1%
usability action	30.2%	28.1%	41.7%
acct security	31.3%	25.4%	43.4%
negative comparison	27.5%	28.5%	44.0%
over enforcement	23.7%	28.6%	47.7%
audience limitation	22.2%	29.9%	47.9%
self harm	28.0%	23.8%	48.2%
bully target	30.1%	21.6%	48.3%
drugs	32.1%	18.7%	49.2%
transparency	23.0%	27.7%	49.3%
data privacy	17.4%	30.5%	52.0%
hate witness	20.8%	25.5%	53.7%
bully witness	20.0%	25.8%	54.2%
nudity	22.5%	23.0%	54.4%
violence	22.6%	22.2%	55.2%
perceived control	19.6%	23.8%	56.5%
political posts	16.3%	24.1%	59.6%
usability passive	14.9%	20.9%	64.2%
misinfo	15.2%	20.1%	64.7%
commerciality	11.8%	21.0%	67.2%
unwanted advances	10.1%	18.4%	71.6%
fake acct 1st	7.9%	14.7%	77.4%

An expanded table with confidence intervals can be found [here](#).



Results by age and gender

Respondents experiencing one or more issues

A little over half of respondents (51.6%) experienced at least one issue in the past 7 days

Each respondent was asked whether they’ve experienced any of five issues, randomly selected out of a set of 22 issues. 51.6% of respondents said they’ve experienced at least one issue. Within age groups, the percentage is highest among those age 16-17 (57.3%), followed by 18-21 (55.7%), and 13-15 (54.1%). The group with the lowest percentage of experiencing one or more issue is age 45+, at 31.2%. (Note: Age was the last question asked, so we only have age data for the 67.5% of respondents who completed the survey.)

	None	1 or more (out of 5)	% of each age group experiencing one or more issues
13-15	5,283	6,215	54.1%
16-17	7,371	9,888	57.3%
18-21	16,033	21,164	56.9%
22-26	12,520	15,713	55.7%
27-34	13,019	13,426	50.8%
35-44	9,978	7,689	43.5%
45 or older	7,877	3,564	31.2%
Prefer not to say	5,647	5,229	48.1%
Total	77,727	82,888	51.6%

Issues by most and least impacted age groups

The 13-15 age group has the highest rates for 8/22 issues, while the 45+ group has the lowest rates of 21/22 issues

The table to the right shows the percentages of respondents who selected “Yes, during the last 7 days” for each of the issues, split by self-reported age. The green cells represent the lowest incidence for that issue across all age groups, and the red cells represent the highest incidence. Numbers that share a numeric subscript are not statistically significantly different from each other at $p < 0.05$. The full breakdown by all age groups is on the next slide.

The 13-15 age group has the most red cells (8)—that is, they have the highest rates of issue experience in 8 of the 22 issues, more than any other age group. By contrast, the oldest age group (45+) has the most green cells (21)—they have the lowest rates of issue experience in 10 of the 22 cells.

	Overall	13-15	45+
commerciality	48.2%	37.7%a	39.6%a,f
misinfo	30.3%	24.4%a	16.1%e
bully witness	28.3%	27.2%a	10.6%e
audience limitation	26.8%	28.2%a,b	15.9%e
usability passive	25.5%	30.1%a	10.1%e
hate witness	25.3%	26.0%a	9.0%e
data privacy	24.4%	17.9%a	14.3%e
perceived control	23.9%	18.4%a	14.2%d
fake acct 1st	21.8%	15.8%a	13.2%e
negative comparison	19.2%	21.4%a	6.8%e
transparency	17.5%	14.9%a	8.4%e
political posts	17.0%	14.3%a,d	8.7%e
nudity	16.3%	19.2%a	6.2%e
over enforcement	14.8%	14.1%a,c	5.9%e
violence	12.8%	12.8%a,b	5.0%d
unwanted advances	11.9%	13.0%a	6.9%d
usability action	10.7%	11.2%a,e	6.6%d
bully target	8.1%	10.8%a	3.4%e
self harm	6.7%	8.4%a	1.9%d
acct security	3.9%	5.2%a	1.3%d
drugs	3.9%	3.7%a,b	1.7%c
impersonation 1st	3.7%	5.5%a	0.3%b

An expanded table with confidence intervals can be found [here](#).

Issues by all age groups

Younger people report higher rates of every issue, with some issues felt more universally than others

In general, the older the respondent, the less likely they are to report experiencing any of the 22 issues measured. All red cells (the age group with the highest rate of experiencing an issue) are within the four youngest age groups.

The gap between the lowest and highest issue rates vary by issue. Bullying witness, for instance, has the widest delta: 31.1% of the 22-26 age group report experiencing it, compared to only 10.6% of the 45+ group, a delta of 20.5 percentage points. Other large deltas include usability passive (delta = 20 points), hate witness (19.5), and misinfo (17.5).

By contrast, certain issues are experienced at similar rates regardless of age. For example, drug sales (delta = 2.1), account security (3.9), and usability action (4.6).

	Overall %	13-15	16-17	18-21	22-26	27-34	35-44	45+	Delta
		Column N %	Column N %	Column N %	Column N %	Column N %	Column N %	Column N %	Highest-lowest
commerciality	48.2%	37.7%a	42.5%b,f,g	47.6%c	51.2%d	50.5%d	44.8%b,e	39.6%a,f	13.50%
misinfo	30.3%	24.4%a	29.2%b	30.0%b,d	33.6%c	31.2%d	24.4%a	16.1%e	17.50%
bully witness	28.3%	27.2%a	29.4%b	29.9%b,c	31.1%c	27.4%a	19.8%d	10.6%e	20.50%
audience limitation	26.8%	28.2%a,b	29.3%a	28.1%a	26.7%b	24.9%c	20.3%d	15.9%e	13.40%
usability passive	25.5%	30.1%a	30.0%a	29.5%a	25.9%b	21.6%c	17.1%d	10.1%e	20.00%
hate witness	25.3%	26.0%a	28.5%b	27.9%b	27.9%b	22.5%c	15.7%d	9.0%e	19.50%
data privacy	24.4%	17.9%a	21.4%b	23.0%c	25.9%d	24.8%d,f	20.2%b	14.3%e	11.60%
perceived control	23.9%	18.4%a	20.0%a	24.8%b	26.7%c	24.7%b,c	20.9%a	14.2%d	12.50%
fake acct 1st	21.8%	15.8%a	21.2%b	24.8%c	23.4%c,d	22.9%b,d	17.4%a	13.2%e	11.60%
negative comparison	19.2%	21.4%a	19.5%b	20.1%a,b	19.4%b	17.5%c	13.4%d	6.8%e	14.60%
transparency	17.5%	14.9%a	19.1%b,f	19.0%b	17.1%c	15.1%a	12.0%d	8.4%e	10.70%
political posts	17.0%	14.3%a,d	16.3%a,b	17.3%b	18.0%b,c	15.2%a,f	12.4%d	8.7%e	9.30%
nudity	16.3%	19.2%a	18.4%a	18.2%a	16.7%b	13.6%c	11.3%d	6.2%e	13.00%
over enforcement	14.8%	14.1%a,c	16.8%b	16.3%b	14.8%a,f	12.9%c	9.1%d	5.9%e	10.90%
violence	12.8%	12.8%a,b	14.3%a	13.2%a	13.1%a	11.5%b	8.2%c	5.0%d	9.30%
unwanted advances	11.9%	13.0%a	14.1%a	14.2%a	11.0%b	8.0%c	7.7%c,d	6.9%d	7.30%
usability action	10.7%	11.2%a,e	11.1%a,e	11.1%a	10.3%a,b	9.6%b	8.5%c	6.6%d	4.60%
bully target	8.1%	10.8%a	9.7%a	8.1%b	7.1%b,c	6.5%c	4.9%d	3.4%e	7.40%
self harm	6.7%	8.4%a	7.2%a,b	7.5%a	5.8%b	3.5%c	2.5%c,d	1.9%d	6.50%
acct security	3.9%	5.2%a	3.2%b,c	4.2%a,b	3.2%c	2.1%d	1.6%d	1.3%d	3.90%
drugs	3.9%	3.7%a,b	3.5%a,b	3.8%a	2.7%a,b,c	3.6%a,b	2.3%b,c	1.7%c	2.10%
impersonation 1st	3.7%	5.5%a	3.5%a	5.0%a	3.1%a	1.2%b	0.7%b	0.3%b	5.20%

An expanded table with confidence intervals can be found [here](#).

Issue rates by age group and gender*

Males and females experience issues at significantly different rates, and those patterns remain consistent across age groups.

Pairs of green cells within each age group represent statistically significant comparisons where females experienced the issues at higher rates than males. For example, in the 13-15 age group, 27.4% of females experienced negative comparison in the last 7 days, compared to only 14.6% of males.

Pairs of yellow cells within each age group represent statistically significant comparisons where males experienced issues at higher rates than females. For example, 14.4% of males in the 13-15 age group reported being the target of bullying in the last 7 days, compared to 8.6% of females.

Interestingly, when an issue is experienced differently by males and females, that difference tends to hold across age groups. The only exception is data privacy concerns, where males report higher rates in the 13-15 and 16-17 age groups, but females report higher rates from age 27 and up.

	13-15		16-17		18-21		22-26		27-34		35-44		45 or older	
	M	F	M	F	M	F	M	F	M	F	M	F	M	F
acct security	5.6%a	4.2%a	3.3%a	2.9%a	3.8%a	3.6%a	3.2%a	3.4%a	1.6%a	2.6%a	1.4%a	1.5%a	1.5%a	1.3%a
audience limitation	26.0%a	31.0%b	26.7%a	33.9%b	27.3%a	31.0%b	24.9%a	30.2%b	24.7%a	25.1%a	22.1%a	21.6%a	18.7%a	14.5%b
bully target	14.4%a	8.6%b	9.8%a	9.6%a	8.7%a	7.9%a	7.5%a	6.0%a	8.1%a	5.0%b	6.7%a	3.6%b	4.7%a	2.4%b
bully witness	27.7%a	27.8%a	28.1%a	30.4%a	28.2%a	32.1%b	29.0%a	32.4%b	25.4%a	28.7%b	18.8%a	21.8%b	11.1%a	11.4%a
commerciality	37.2%a	39.4%a	41.2%a	42.4%a	47.7%a	45.4%a	51.0%a	50.7%a	51.7%a	49.2%a	48.3%a	42.8%b	40.0%a	39.9%a
data privacy	16.4%a	20.6%b	19.2%a	22.1%b	22.9%a	22.7%a	26.4%a	25.0%a	27.3%a	22.8%b	23.0%a	19.4%b	16.3%a	12.1%b
drugs	3.1%a	4.4%a	4.1%a	2.7%a	4.9%a	3.5%a	2.3%a	2.4%a	5.6%a	2.1%b	1.8%a	0.9%a	1.5%a	1.7%a
fake acct 1st	20.1%a	14.7%b	23.9%a	19.2%b	25.5%a	22.4%b	23.7%a	23.8%a	23.3%a	22.6%a	23.6%a	14.2%b	15.6%a	12.6%a
hate witness	23.7%a	27.8%a	25.9%a	29.4%b	26.1%a	28.7%b	25.9%a	28.6%b	21.7%a	23.0%a	15.9%a	16.2%a	9.4%a	9.4%a
impersonation 1st	4.3%a	1.2%a	2.9%a	4.4%a	5.2%a	3.6%a	4.0%a	2.9%a	0.4%a	1.2%a	1.6%a	0.5%a	0.3%a	0.5%a
misinfo	28.8%a	22.3%b	33.2%a	26.1%b	32.6%a	28.3%b	36.7%a	30.5%b	35.8%a	27.3%b	28.9%a	22.7%b	17.8%a	16.0%a
negative comparison	14.6%a	27.4%b	12.8%a	25.2%b	14.9%a	24.3%b	15.0%a	22.5%b	13.8%a	19.2%b	12.1%a	14.4%b	6.5%a	6.3%a
nudity	20.7%a	17.7%a	20.0%a	17.7%a	17.8%a	16.7%a	17.4%a	16.1%a	13.4%a	12.9%a	11.3%a	10.7%a	6.5%a	6.3%a
over enforcement	16.3%a	15.2%a	17.9%a	14.5%b	17.7%a	14.9%b	14.0%a	15.1%a	13.2%a	11.5%b	11.0%a	8.4%b	9.3%a	3.7%b
perceived control	22.2%a	19.7%a	21.0%a	19.3%a	24.9%a	21.4%b	29.7%a	24.2%b	24.8%a	23.7%a	21.0%a	20.1%a	16.3%a	15.0%a
political posts	19.1%a	13.8%b	17.0%a	14.6%a	20.6%a	15.3%b	20.4%a	16.1%b	17.9%a	14.2%b	14.4%a	10.3%b	11.3%a	9.2%a
self harm	8.2%a	9.5%a	6.3%a	6.1%a	7.5%a	7.2%a	6.7%a	5.6%a	4.6%a	2.9%a	4.1%a	1.4%b	2.0%a	1.0%a
transparency	19.0%a	13.5%b	21.2%a	19.0%a	20.8%a	17.5%b	18.1%a	16.3%b	17.1%a	14.2%b	15.3%a	10.7%b	11.2%a	7.3%b
unwanted advances	14.4%a	11.6%b	17.1%a	12.2%b	15.8%a	12.6%b	13.1%a	9.1%b	10.6%a	6.4%b	9.6%a	6.5%b	11.0%a	4.2%b
usability action	8.6%a	12.6%b	10.2%a	11.0%a	9.5%a	12.1%b	9.5%a	11.5%b	8.9%a	10.1%a	8.1%a	9.6%a	7.5%a	6.4%a
usability passive	24.9%a	34.5%b	29.1%a	31.5%a	27.5%a	31.6%b	23.7%a	28.0%b	20.4%a	23.2%b	16.2%a	17.8%a	9.8%a	10.5%a
violence	14.1%a	11.5%a	14.4%a	13.9%a	12.2%a	12.8%a	12.9%a	11.7%a	10.9%a	12.0%a	8.1%a	8.1%a	6.2%a	4.5%a

*Gender is measured with log data. 37.3% of the sample was labeled male, 35.1% female, 27.2% unknown, and 0.5% custom. Unknown and custom were excluded from this analysis.

An expanded table with confidence intervals can be found [here](#).

Bullying and known antagonists, by gender

Females respondents are more likely than males to know the antagonist offline when the target of bullying, but less likely to know the antagonist offline when witnessing bullying.

The percentage of females responding that they know the antagonist offline/ in real life for bullying target is 17.4%, versus 11.8% for males. This difference is statistically significant, measured via a 2x2 chi square test (the ‘I know them, but only online’ and ‘I don’t know them’ response options were combined for this analysis).

Interestingly, the pattern is reversed for bullying witness. The percentage of females responding that they know the antagonist offline/ in real life for bullying witness is 5.5%, versus 7.2% for males. This difference is also statistically significant.

These two patterns are particularly pronounced in the 13-15 age group (bottom table), although the pattern is found in most age groups. Because of small sample sizes when cutting by age, though, separate statistical tests were not run for those subsets.

Known antagonist by gender

		male			female		
		Offline	Online	DK	Offline	Online	DK
Bullying target	Row %	11.8%	16.6%	71.6%	17.4%	13.0%	69.5%
	95% Lower CI	9.3%	13.7%	67.8%	14.3%	10.4%	65.5%
	95% Upper CI	14.5%	19.7%	75.1%	20.8%	16.1%	73.4%
	n	69	97	420	92	69	368
Bullying witness	Row %	7.2%	15.5%	77.3%	5.5%	21.1%	73.4%
	95% Lower CI	6.3%	14.2%	75.7%	4.8%	19.8%	72.0%
	95% Upper CI	8.2%	16.9%	78.7%	6.3%	22.3%	74.8%
	n	211	455	2266	219	836	2916

Known antagonist by gender, ages 13-15 only

		male			female		
		Offline	Online	DK	Offline	Online	DK
Bullying target	Row %	17.3%	24.6%	58.1%	29.5%	23.3%	47.2%
	95% Lower CI	8.7%	14.5%	45.1%	17.4%	12.6%	33.5%
	95% Upper CI	28.7%	37.2%	71.0%	42.2%	35.5%	60.8%
	n	9	13	31	14	11	23
Bullying witness	Row %	8.4%	20.1%	71.5%	5.4%	17.3%	77.3%
	95% Lower CI	4.3%	13.4%	62.9%	3.0%	13.0%	71.9%
	95% Upper CI	14.1%	27.4%	78.8%	8.5%	22.4%	82.3%
	n	10	24	87	13	43	192



Results by creator segmentation

Creator segmentation

5.41% of survey respondents are classified as creators, based on ig_creator_segment.

Because there are too few emerging (n=51) and established (n=9) respondents in the dataset to analyze those categories separately, they were combined with aspiring (n=698) respondents into an ‘aspiring+’ group for further analyses. The former pre-experimenting category was added to the not creator category.

		IG creator segment				
		Not creator	Experimenting	Aspiring	Emerging	Established
feed_media_7d	Mean	1	2	4	8	6
	Median	0	0	1	3	4
	n	225,143	12,128	698	51	9
story_media_7d	Mean	7	19	35	51	78
	Median	1	11	21	29	60
	n	225,143	12,128	698	51	9
reels_media_7d	Mean	0	0	1	1	0
	Median	0	0	0	0	0
	n	225,143	12,128	698	51	9
followers	Mean	514	2,110	22,843	266,696	1,807,007
	Median	182	1,512	15,683	178,042	1,556,500
	n	225,143	12,128	698	51	9

Issues by creator segmentation

Most issues are experienced more often by larger creators, but the gap varies by issue

In this table, issues are ranked by the issue rates of the aspiring+ creator group (a combination of the aspiring, emerging, and established segments). Red/green cells are the highest/lowest value in each row. 12 of the 22 issues are experienced most often by the aspiring+ group, while nearly all (21 of 22) issues are experienced most often by either the large or experimenting groups. By contrast, 21 of 22 issues are experienced the *least often* by respondents without a creator status.

Similar to age, not all issues have large gaps in experience rates across creator status. Not surprisingly, audience limitation (“*Have you felt that not enough people see the things you share on Instagram?*”) is both the most frequent issue for aspiring+ creators, as well as the issue with the largest gap: 53.5% of aspiring+ creators experienced this issue in the past 7 days, compared to only 25.7% of non-creators, a delta of 27.8 percentage points. Bullying target has the second highest delta (18.3), followed by fake account (15.9).

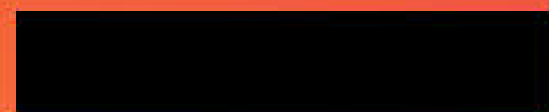
By contrast, issues such as drugs (1.2), account security (1.4), data privacy (1.7), and violence (2.0) are experienced similarly regardless of creator status.

	None	Experimenting	Aspiring+	Delta
Issue	Column N %	Column N %	Column N %	High-low
audience limitation	25.7%c	45.5%b	53.5%a	27.8%
bully witness	28.0%b	34.0%a	37.7%a	9.7%
misinfo	30.2%b	31.9%a	37.5%a	6.1%
fake acct 1st	21.4%b	29.2%a	37.3%a	15.9%
commerciality	48.2%b	48.0%a,b	36.3%a	11.9%
usability passive	25.0%b	33.6%a	32.1%a	8.6%
hate witness	25.1%b	30.0%a	28.4%a,b	4.9%
bully target	7.8%c	12.1%b	26.1%a	18.3%
data privacy	24.4%a	24.7%a	26.1%a	1.7%
perceived control	23.7%b	26.4%a	24.7%a,b	2.7%
nudity	16.2%a	17.4%a	21.3%a	5.1%
political posts	16.8%b	21.1%a	20.4%a,b	4.3%
unwanted advances	11.7%b	15.2%a	19.2%a	7.5%
negative comparison	19.0%b	21.9%a	17.9%a,b	4.0%
transparency	17.3%b	20.4%a	16.6%a,b	3.8%
usability action	10.4%b	15.6%a	16.2%a	5.8%
over enforcement	14.6%b	16.8%a	15.3%a,b	2.2%
violence	12.7%b	14.7%a	14.7%a,b	2.0%
impersonation 1st	3.7%b	3.2%b	14.7%a	11.5%
self harm	6.6%b	8.9%a	12.1%a,b	5.5%
drugs	3.9%a	5.1%a	4.0%a	1.2%
acct security	3.8%b	5.2%a	3.8%a,b	1.4%

*Aspiring, emerging, and established creator segments combined

**Cells in each row not sharing a subscript are statistically significant from each other, p < 0.05

An expanded table with confidence intervals can be found [here](#).



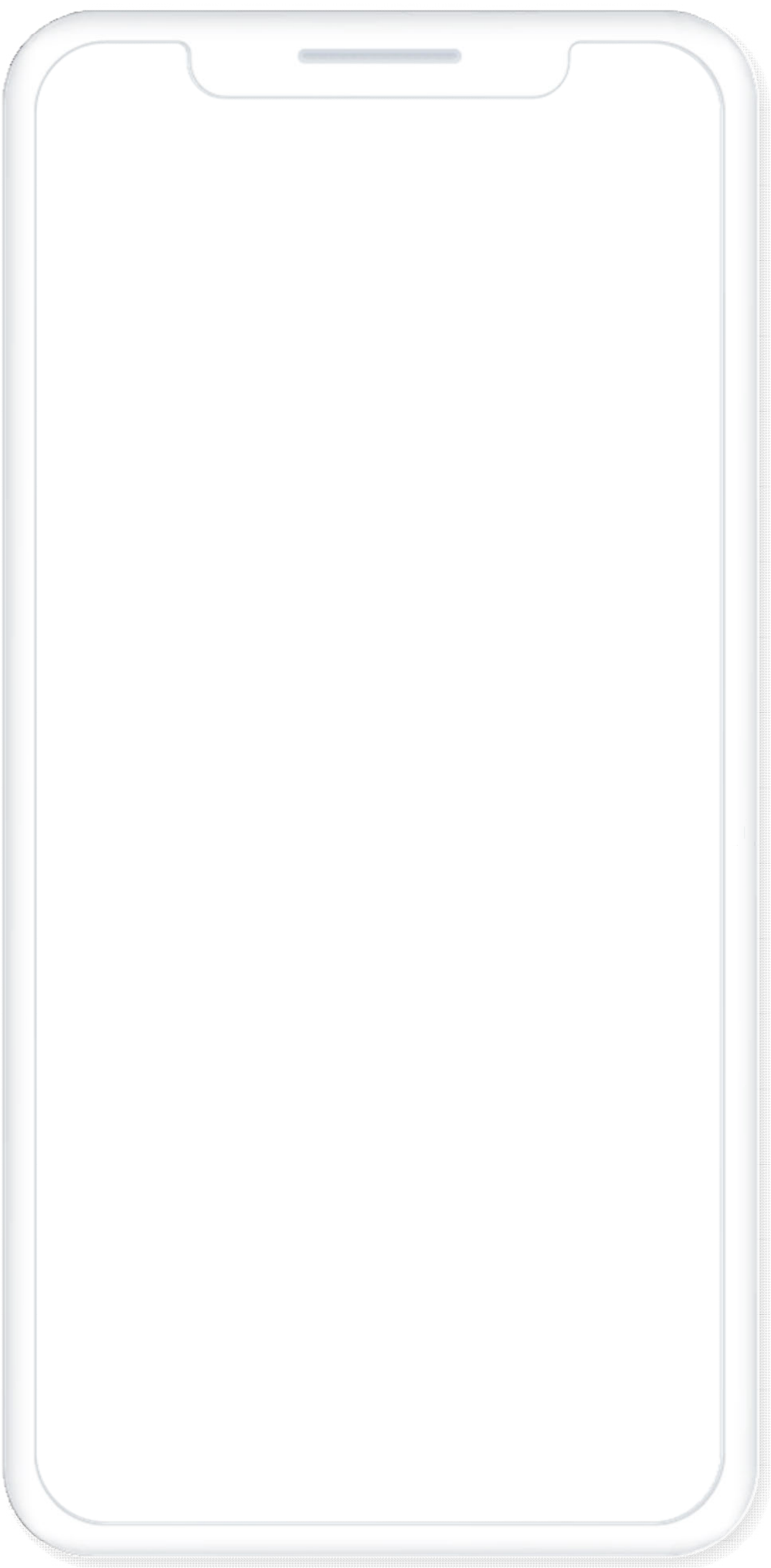
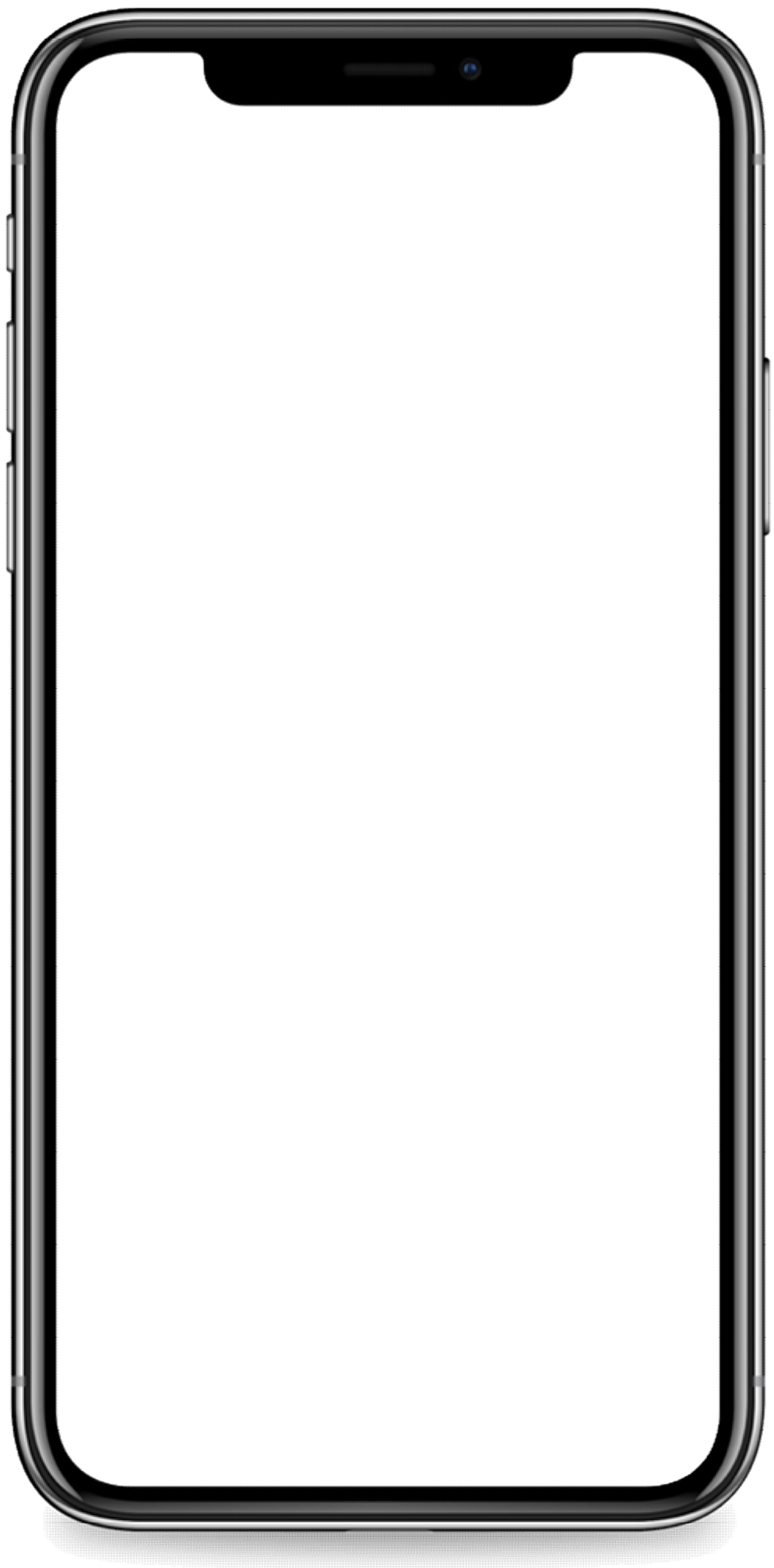
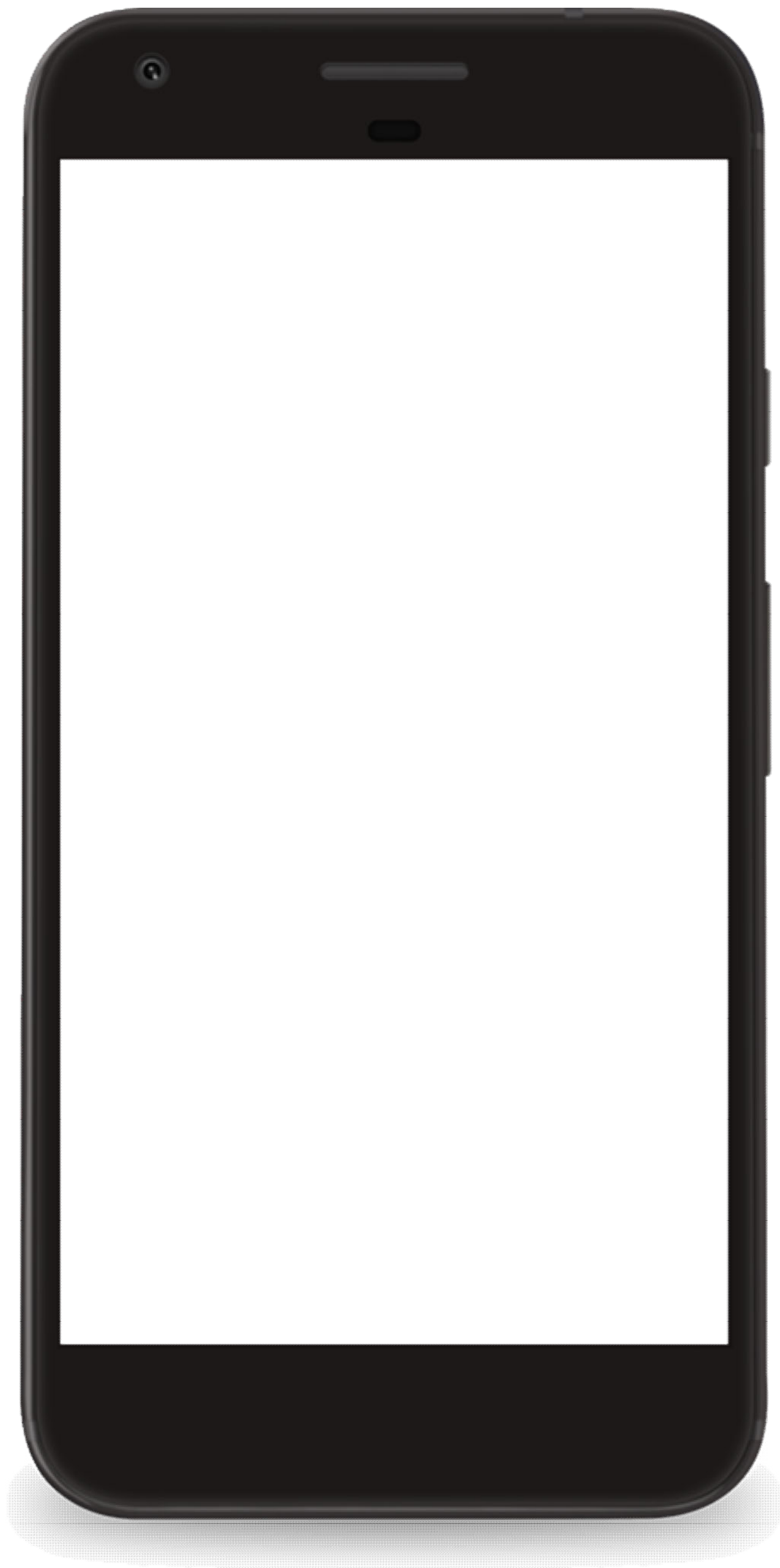






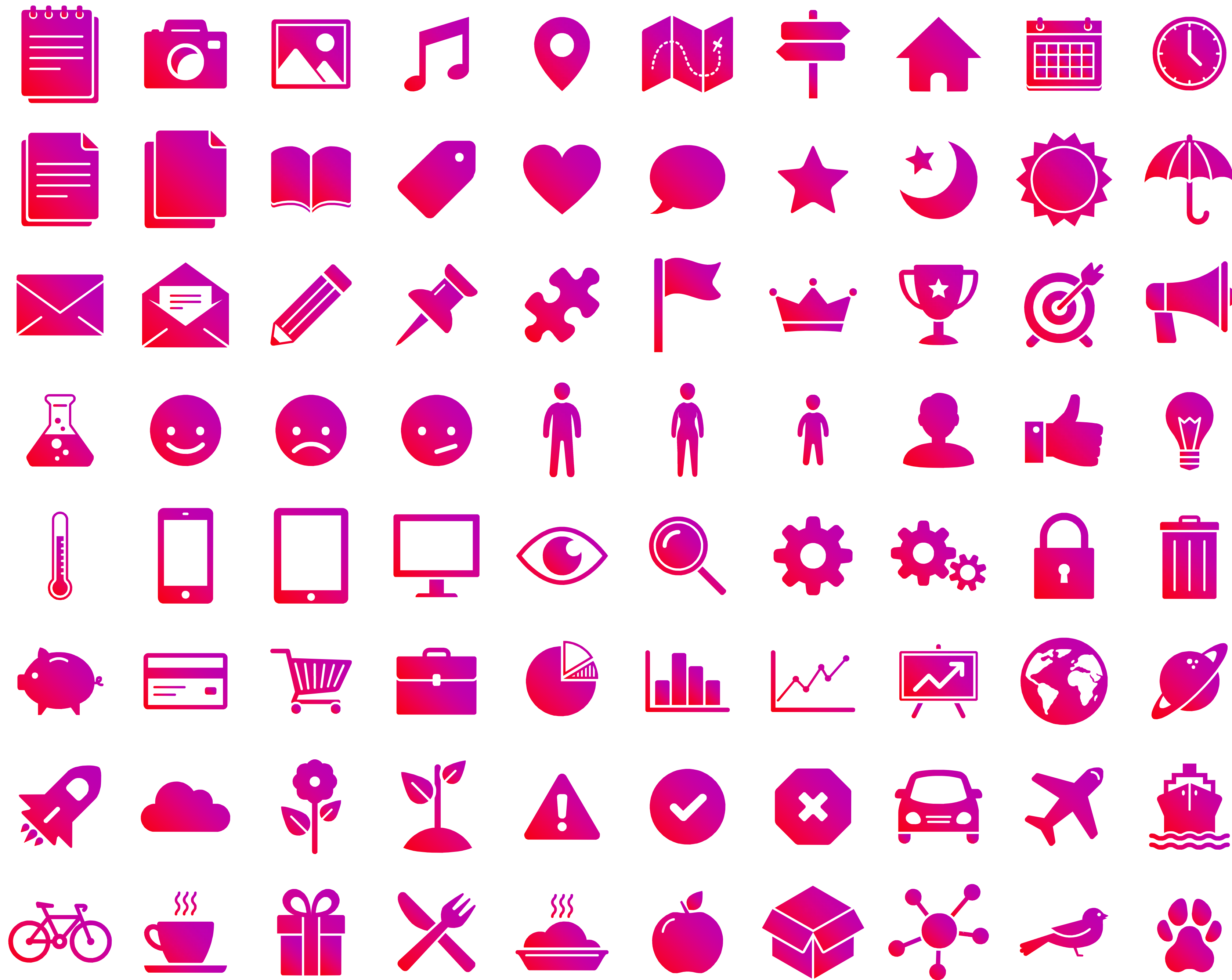






facebook





Editable Icons

Just copy and paste and
change color or lines
accordingly